BANCA D'ITALIA

# Temi di discussione

## del Servizio Studi

## Gaussian inference on certain long-range dependent volatility models

by Paolo Zaffaroni

Number 472 - June 2003

# GAUSSIAN INFERENCE ON CERTAIN LONG-RANGE DEPENDENT VOLATILITY MODELS

**Paolo Zaffaroni** *

**Abstract**

For a class of long memory volatility models, we establish the asymptotic distribution theory of the Gaussian estimator and the Lagrange multiplier test. Both the case of estimation of martingale difference and ARMA levels are considered. A Montecarlo exercise is presented to assess the small sample properties of the Gaussian estimator and the Lagrange multiplier test. An empirical application, using foreign exchange rates and stock index returns, suggests the potential of these models to capture the dynamic features of the data.
JEL classification: C22, C51, G12
Keywords: volatility model; nonlinear moving average model; long memory; Whittle estimation; asymptotic distribution theory.

**Contents**



*Bank of Italy, Economic Research Department.

# 1 Introduction*

The importance of modelling volatility lies in the dependence of any financial investment decision on the forecasts of asset risks and returns, formalized in asset pricing theory following the Sharpe (1964) and Lintner (1965) capital asset pricing model. Moreover, a deep understanding of the data generating process underlying a time series is a compulsory step towards forecasting it and, when dealing with financial assets, is crucial for a successful implementation of any option-pricing model. Indeed, since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), a great deal of research has focused on modelling the volatility of financial asset returns as they feature very complicated dynamics, synthesized in a number of well-known stylized facts, dynamic conditional heteroskedasticity in particular. Estimation of these models is usually based on a maximum likelihood estimator (MLE) approach, mainly using the assumption of Gaussianity of rescaled innovations. This is easily implementable but derivation of the statistical properties of MLE, pseudo MLE (PMLE) when the distributional assumptions are violated, appears a very difficult task because of the nonlinear feature of the model. Indeed, consistency and asymptotic normality have been established for the Gaussian PMLE of the basic ARCH($p$) model (see Weiss (1986)) and for the generalized ARCH$(1,1)$ (GARCH$(1,1)$) only (see Lee and Hansen (1994) and Lumsdaine (1996)), despite the innumerable developments of ARCH-type volatility modelling. (See e.g. Bollerlsev, Chou, and Kroner (1992) or Bollerslev, Engle, and Nelson (1995) for complete surveys.)

In this paper we consider an alternative way of modelling changing volatility by means of a family of nonlinear moving average (nonlinear MA) models, introduced by Robinson and Zaffaroni (1997),

$$x_t = \mu + u_t, \qquad t \in \mathbb{Z}, \tag{1}$$

with

$$u_t = \epsilon_t h_{t-1}, \quad h_{t-1} = \rho + \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i}, \quad \sum_{i=1}^{\infty} \alpha_i^2 < \infty, \tag{2}$$

where the $x_t$ represent the observed time series and the $\epsilon_t$ the unobservable independent identically distributed (*i.i.d.*) zero mean innovations. The constant $\mu$ defines the (conditionally constant) first moment of the $x_t$ whereas $\rho$ and the $\alpha_i$ characterize the memory of squares and higher-order statistical properties, as discussed below. Robinson (2001) shows that for Gaussian $\epsilon_t$ the nonlinear MA model (2) is nested within a large class of volatility models, which includes the standard stochastic volatility (SV) model of Taylor

*Reprinted from *Journal of Econometrics*, copyright (2003), with permission from Elsevier Science.

(1986) and a simple case of the exponential GARCH (EGARCH) model of Nelson (1991).

The main contribution of this paper is to establish the asymptotic statistical properties of the frequency domain Gaussian estimator in the sense of Whittle (1962), for the nonlinear MA (1)-(2), employing the squares as the observable. We focus on parameterizations for the model that allow long memory autocorrelated squares. This choice is motivated on both theoretical and empirical grounds. In fact, although not formally established, it is likely that generalizations of previous work on Gaussian estimation and method-of-moment estimation, along the lines of Robinson (1977,1978), would yield asymptotically correct statistical inference for short memory parameterizations of the squares. On the other hand, relatively recent empirical research on asset returns volatility suggests that the effect of shocks on the conditional variance is very persistent though eventually absorbed with time (see Gallant, Rossi, and Tauchen (1993)), ruling out integrated GARCH behaviour. That is, sample autocorrelations in squared returns tend to decline very slowly in contrast with the fast exponential decay implied by standard ARCH-type models. This is, in turn, consistent with the notion of long memory when theoretical autocovariances are not summable, or when, alternatively, the power spectrum is unbounded at zero frequency.

With respect to ARCH-type models various alternatives have been proposed to account for this aspect of asset returns dynamics. In order to develop a test for no-ARCH Robinson (1991) introduced the ARCH($\infty$), a possibly long memory generalized ARCH model. Baillie, Bollerslev, and Mikkelsen (1996) considered a particular case of the ARCH($\infty$) denominated as fractionally integrated GARCH (FIGARCH), and proposed a time domain Gaussian PMLE estimation approach. Note that ARCH($\infty$) is a nonlinear autoregressive model and cannot be nested within the Robinson (2001) class of nonlinear moving average models. Nelson (1991), introducing the EGARCH, mentioned the possibility of long memory parameterizations explicitly developed in Bollerlsev and Mikkelsen (1996, the fractionally integrated EGARCH). These long memory volatility models are characterized by the lack of asymptotic distribution theory for the PMLE, as well as for any other estimator, and the available asymptotic results for estimation of short memory ARCH-type models and linear long memory processes are not readily extendable.

By contrast, in this paper we provide a formal framework to assess whether long memory volatility models represent a valid alternative to short memory ones.

Making distributional assumptions on the innovations, the nonlinear MA could be estimated by (exact) MLE. However, asymptotic properties of the MLE would depend on invertibility of the model and the invertibility theory of linear processes is not immediately extendable to nonlinear cases, as noted in Granger and Andersen (1978). Furthermore, for the nonlinear MA, likelihood and scores would only be computable recursively and expensively, which would render asymptotic theory of exact MLE extremely cumbersome to establish. Moreover, given the dynamic structure of the model, the likelihood necessarily depends on the initial conditions which will probably have non-negligible effects in finite samples, especially within a long memory parameterization. More importantly, the asymptotic properties of the MLE may not be robust with respect to deviations from the postulated distributional assumption, as often happens for nonlinear econometric models.

These arguments do not apply to the Gaussian estimator considered here, partly owing to its frequency domain specification. We establish that the Gaussian estimator is strongly $\sqrt{T}$-consistent with a normal asymptotic distribution, $T$ denoting the sample size. This result does not follow directly from the asymptotic results on ARCH estimation nor from the Gaussian estimation theory for linear long memory models (see e.g. Giraitis and Surgailis (1990), Hosoya (1997)), but depends upon a central limit theorem (CLT) for quadratic forms in nonlinear long memory variates established in this paper. Secondly, we develop a pseudo Gaussian Lagrange multiplier (LM) testing procedure designed to detect dynamic conditional heteroskedasticity that uses the nonlinear MA as a class of alternatives. Despite the long memory feature of these alternative hypotheses, the usual asymptotic chi-square distribution is obtained under the null hypothesis of homoskedasticity.

The paper is organized as follows. In section 2 we briefly recall the low-order statistical properties of nonlinear MA: memory, kurtosis and 'leverage' effect. Estimation of the model is discussed in section 3, where we introduce the Gaussian estimator and the LM multiplier test and establish their asymptotic statistical properties. The martingale difference assumption for (demeaned) levels $x_t$ implied by (1) and (2) makes it possible to focus exclusively on the volatility dynamics of the nonlinear MA model. However, it represents at best an approximation for practical use of the model. A finite-order autoregressive moving average (ARMA) model represents a simple yet appealing parameterization of the conditional mean of the observable $x_t$. In section 4, we present the asymptotic properties of the Whittle estimator of the parameters of ARMA models with innovations $u_t$ described by the nonlinear MA (2). Preliminary estimation of the nonlinear MA model is not required. Next, we present the asymptotic properties of the Gaussian estimator of the nonlinear MA based on the estimated ARMA residuals. Section 5 reports the result of Montecarlo exercises to assess the small-sample performance of the Gaussian estimator and of the LM test. In section 6 we present an empirical application based on observed time series of stock index and foreign exchange rate returns. The nonlinear MA and its likely main competitor, the ARCH($\infty$), are compared in terms of goodness-of-fit using the Gaussian estimator. Concluding remarks are in section 7. The results are formally stated in theorems the proofs of which are reported in the mathematical appendixes, together with a number of technical lemmas.

# 2 Low-order statistical properties of the nonlinear moving average model

Let the innovation $\epsilon_t$ in (2) satisfy the following assumption.
**Assumption A.** *The process* $\{\epsilon_t\}$ *is i.i.d. with* $E(\epsilon_t^4) < \infty$ *and*

$$\begin{aligned} E(\epsilon_t) &= 0, \\ E(\epsilon_t^2) &= \sigma^2,\ 0 < \sigma^2 < \infty\,, \\ E(\epsilon_t^3) &= 0. \end{aligned}$$

*Let* $\kappa$ *denote the fourth-order cumulant of the* $\epsilon_t$*, yielding* $E(\epsilon_t^4) = \kappa + 3\sigma^4$.

We do not normalize the innovation variance $\sigma^2$ to be equal to 1. More importantly, we do not constrain the magnitude of the fourth moment and of the fourth-order cumulant. Assumption $A$ is satisfied, for instance, for Student's $t$ $\epsilon_t$ with $n$ degrees of freedom whenever $n > 4$. In this case $\kappa = n^2 B(\frac{5}{2}, \frac{n-4}{2})/B(\frac{1}{2}, \frac{n}{2}) - 3(\frac{n}{n-2})^2$ with $\kappa > 0$ where $B(\cdot,\cdot)$ indicates the Beta function. When the $\epsilon_t$ are Gaussian then $\kappa = 0$ whereas $\kappa < 0$ for $\epsilon_t$ uniformly distributed over $[-a, a]$ for some finite $a > 0$. Note that, in general, bounded fourth moment and non degeneracy of the $\epsilon_t$ ($0 < \mathrm{var}(\epsilon_t^2) < \infty$) implies $\kappa > -2\sigma^4$. When Assumption $A$ holds, the demeaned $x_t$ are martingale difference but not independent as the squares $y_t = x_t^2$ are autocorrelated (cf. Robinson and Zaffaroni (1997)).

A frequency domain characterization of the memory properties of the $x_t$ and the $y_t$ is central to the paper, in particular with respect to the estimation part (cf. section 3). For this purpose, let $\alpha(\lambda)\,(-\pi \le \lambda < \pi)$ be the transfer function of the sequence $\{\alpha_i\}$, implying

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha(\omega) e^{i\omega u} d\omega = \alpha_u,\ \ u \ge 1. \tag{3}$$

Set

$$\beta(\lambda) = 2\mathcal{R}e(\alpha(\lambda)),\ \ -\pi \le \lambda < \pi, \tag{4}$$

$\mathcal{R}e(\cdot)$ denoting the real part of its argument and let $f_{zz}(\lambda)\,(-\pi \le \lambda < \pi)$ be the power spectrum (when it exists) of a weakly stationary process $\{z_t\}$. Hereafter, set $\delta_{ab}(u) = \sum_{i=1}^{\infty} a_i b_{i+u}\,(u = 0, \pm1, \ldots)$ for any square summable sequences $\{a_i\}, \{b_i\}$.

**Theorem 1** *Under Assumption A, for $-\pi \leq \omega < \pi$,*

$$
\begin{aligned}
f_{xx}(\omega) &= \frac{\sigma^2}{2\pi}(\rho^2 + \sigma^2 \delta_{\alpha\alpha}(0)), \\
f_{yy}(\omega) &= \frac{(2\sigma^4 + \kappa)(\rho^2 + \sigma^2 \delta_{\alpha\alpha}(0))}{(2\pi)^2} \int_{-\pi}^{\pi} \beta(\lambda)\beta(\omega - \lambda) d\lambda + \frac{4\sigma^4 \rho^2 \mu}{2\pi} \beta(\omega) \\
&+ \frac{2\sigma^8}{(2\pi)^2} \int_{-\pi}^{\pi} |\alpha(\lambda)|^2 |\alpha(\omega - \lambda)|^2 \, d\lambda + \frac{4\sigma^6 \mu}{(2\pi)^2} \int_{-\pi}^{\pi} \beta(\lambda) \, |\alpha(\omega - \lambda)|^2 \, d\lambda \\
&+ \frac{4\sigma^6 \rho^2}{2\pi} \, |\alpha(\omega)|^2 + \frac{\sigma^4 \kappa}{(2\pi)^3} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \alpha(\lambda)\alpha(\theta)\alpha(\omega - \theta)\alpha(-\omega - \lambda) d\lambda d\theta \\
&+ \frac{\nu_y}{2\pi},
\end{aligned}
$$

*where $\nu_y$ is a strictly positive constant.*

**Proof:** see appendix A.

The power spectrum of the $y_t$ is a linear combination of convolutions of the $|\alpha(\omega)|^2$, $\beta(\omega)$, given the nonlinear transformation involved $y_t = x_t^2$ (see Hannan (1970, p.82)).

Alternatively, under Assumption $A$ the $y_t$ have autocorrelation function (ACF)

$$
\begin{aligned}
\gamma_{yy}(l) &= 2\sigma^8 \alpha_{|l|}^2 \delta_{\alpha\alpha}(0) + 2\sigma^8 \delta_{\alpha\alpha}^2(l) + 4\rho^2 \sigma^6 \delta_{\alpha\alpha}(l) + \sigma^4 \kappa (\delta_{\alpha\alpha}(0) \alpha_{|l|}^2 + \delta_{\alpha^2\alpha^2}(l)) \\
&+ 4\sigma^4 \mu \alpha_{|l|} \left[\rho^2 + \sigma^2 \delta_{\alpha\alpha}(l)\right] + 2\rho^2 \sigma^6 \alpha_{|l|}^2 + \nu_y \delta(l, 0), \;\; l = 0, \pm 1, ... \qquad (5)
\end{aligned}
$$

where, henceforth we set $\gamma_{zu}(l) = \mathrm{cov}(z_t, u_{t+l})$, $l = 0, \pm 1, \ldots$, for stationary processes $\{z_t,\, u_t\}$ and by $\delta(u, v)$ the Kronecker delta. Robinson and Zaffaroni (1997), hereafter RZ, establish (5) for the case $\rho = \mu = 0$. We skip details for simplicity's sake but, in the proof of Theorem 1, sketch the derivation of (5) in the more general case here considered.

By (5) and square summability of the $\alpha_i$ (see RZ), as $l \to \infty$,

$$
\gamma_{yy}(l) \sim \begin{cases} 2\sigma^8 \delta_{\alpha\alpha}^2(l), & \rho = 0, \\ 4\rho^2 \sigma^6 \delta_{\alpha\alpha}(l), & \rho \neq 0, \end{cases} \qquad (6)
$$

where $\sim$ denotes asymptotic equivalence: $a(x) \sim b(x)$ as $x \to x_0$ when $a(x)/b(x) \to 1$ as $x \to x_0$. When

$$
\alpha_j \sim k \, j^{d-1} \quad \text{as} \quad j \to \infty
$$

with $0 < d < 1/2$, $0 <| k |< \infty$, the $y_t$ display an hyperbolically decaying ACF, exhibiting long memory when $\rho \neq 0$ or for $d > 1/4$ when $\rho = 0$ (cf. RZ, section 4). In particular, when $\rho \neq 0$

$$
\gamma_{yy}(l) \sim k \, l^{2d-1} \quad \text{as} \quad l \to \infty, \qquad (7)
$$

and both long memory and covariance stationarity are achieved for $0 < d < 1/2$.

As a by-product of the nonlinearity, the nonlinear MA exhibits fatter tails than the Gaussian case, thus matching another well-known stylized fact of asset return dynamics, since Fama (1965). Under Assumption $A$, the model kurtosis is

$$\begin{aligned}\frac{E(x_t-\mu)^4}{\gamma_{xx}^2(0)} &= 3+\frac{6}{(\bar{\rho}^2+\delta_{\alpha\alpha}(0))^2}(\delta_{\alpha\alpha}^2(0)+2\bar{\rho}^2\delta_{\alpha\alpha}(0)) \\ &+\frac{\bar{\kappa}}{(\bar{\rho}^2+\delta_{\alpha\alpha}(0))^2}\left(\bar{\rho}^4+6\bar{\rho}^2\delta_{\alpha\alpha}(0)+3\delta_{\alpha\alpha}^2(0)+(3+\bar{\kappa})\delta_{\alpha^2\alpha^2}(0)\right),\end{aligned}$$

setting $\bar{\rho}=\rho/\sigma$, $\bar{\kappa}=\kappa/\sigma^4$. The second term on the right-hand side expresses the deviation from the Gaussian kurtosis case due to the nonlinearity of the model whereas the third term depends on the fourth-order cumulant. It will be zero only for $\kappa = 0$ but it can be arbitrarily large, depending on the tail behaviour of the marginal distribution of the $\epsilon_t$.

Since Black (1976), another feature of asset return empirical distributions is the so-called 'leverage' effect, i.e. an asymmetric response of future ($l$-period ahead) volatility to past and present shocks implied by a negative value of $\mathrm{cov}(x_t, y_{t+l})$. Insights into this dynamic asymmetry can be obtained generalizing $h_t$ in (2) by

$$\tilde{h}_t=\rho+\sum_{i=1}^{\infty}\alpha_i g(\epsilon_{t-i}), \tag{8}$$

where $g(\epsilon_t)=\vartheta\epsilon_t+\gamma(|\epsilon_t|-E\,|\epsilon_t|)$ for some constant real $\vartheta,\gamma$. The specification of the $g(\cdot)$ function, known as the 'news impact curve' in volatility literature, was introduced by Nelson (1991). Under Assumption $A$, setting (for simplicity) $\mu = 0$ and $E(\epsilon_t|\epsilon_t|) = 0$, e.g. implied by symmetric $\epsilon_t$, one obtains for $l > 0$

$$\mathrm{cov}(x_t,y_{t+l})=2\sigma^4\vartheta\alpha_l\left((\sigma^2(\vartheta^2+\gamma^2)-\gamma^2E^2|\epsilon_t|)\delta_{\alpha\alpha}(l)+\rho^2\right).$$

It follows that the crucial (necessary) condition to account for the 'leverage' effect is $\vartheta \neq 0$. The interplay of $\vartheta$ and $\gamma$ would impart great flexibility in modelling asymmetry expressed by shifts and rotations in the 'news impact curve' (see e.g. Hentschel (1995)). We decided to focus on the particular yet asymmetric case $h_t$ (cf. (2)), setting $\vartheta = 1, \gamma = 0$ in $\tilde{h}_t$. In fact, both the non-differentiability of the absolute-value function in $g(\cdot)$ and the possibility of leaving $\vartheta$, $\gamma$ unrestricted are likely to induce a great deal of technical difficulties when establishing the asymptotics for the class of estimators considered in this paper (cf. section 3). The 'leverage' effect, at lag $L$, arises whenever $\alpha_L < 0$ independently of the value of $\rho$. The nonlinear MA is also apt to display negative 'simultaneous' correlation between asset return and asset volatility, as observed in Campbell and Hentschel (1990).

In contrast to nonlinear autoregressive models (e.g. ARCH-type models), derivation of the (low-order) statistical properties of the nonlinear MA appears straightforward. Note, however, that this is not achieved by introducing a second latent sequence of innovations, as for example for SV models. Thus, the usual difficulties arising with the estimation and filtering of SV models are avoided in our nonlinear MA framework. We now discuss estimation and testing of the model.

# 3 Estimation and testing

## 3.1 Parameterization

Within this section, let us set $\mu = 0$ for the sake of simplicity. The more complicated case of observable $x_t$ with time-varying conditional mean will be described in section 4. Let us introduce a sequence of functions $\alpha_i(\theta)$ $(i \geq 1)$ of a $p \times 1$ vector $\theta \in \Theta \subset R^p$. Within the set-up of the previous sections, this implies introducing new functions $\alpha(\lambda; \theta)$ and $\beta(\lambda; \theta)$ of the frequency $\lambda\,(-\pi \leq \lambda < \pi)$ and $\theta$. Replace $\alpha(\lambda)$ and $\beta(\lambda)$ in $f_{yy}(\lambda)$ (cf. Theorem 1) with $\alpha(\lambda; \theta)$ and $\beta(\lambda; \theta)$. Recalling $\bar{\kappa} = \kappa/\sigma^4$, $\bar{\rho}^2 = \rho^2/\sigma^2$, set $\psi = (\bar{\kappa}, \theta')'$ and $\bar{\rho}^2 = 1$ for model identification. This yields $f_{yy}(\lambda; \psi, \sigma^2)$ for $\psi \in \Psi \subset R^{p+1}$, $\Psi$ denoting the parameter space. It is convenient to express $f_{yy}(\lambda; \psi, \sigma^2)$ as $\sigma^8 h_y(\lambda; \psi)$ with $h_y(\cdot; \psi) = f_{yy}(\cdot; \psi, 1)$ because, as shown below, rescaling $\kappa$ and $\rho^2$ by suitable powers of $\sigma^2$ allows factorizing and hence concentrating $\sigma^2$ out. In the same way, replacing the $\alpha_i$ with the $\alpha_i(\theta)$ in (5), we define the sequence $\gamma_{yy}(l; \psi, \sigma^2)$, $l = 0, \pm 1, \ldots$, which can be re-written as $\gamma_{yy}(l; \psi, \sigma^2) = \sigma^8 \bar{\gamma}_{yy}(l; \psi)$. Denoting by $\psi_0$ the true value for $\psi$, with $\psi_0 \in \Psi$, it follows that, for any $-\pi \leq \lambda < \pi$ and $l = 0, \pm 1, \ldots$,

$$f_{yy}(\lambda) = \sigma^8 h_y(\lambda; \psi_0) = \sigma^8 h_y(\lambda), \;\; \gamma_{yy}(l) = \sigma^8 \bar{\gamma}_{yy}(l; \psi_0), \;\; \alpha_l = \alpha_l(\theta_0).$$

An example of a parsimonious parameterization, adopted in our empirical application below (cf. section 6), is obtained by choosing the $\alpha_i(\theta)$ as the MA($\infty$) coefficients of the autoregressive fractional moving average model of order $p, d, q$ (ARFIMA$(p, d, q)$) of Granger and Joyeux (1980), for integers $p, q \geq 0$ and real $\mid d \mid < 1/2$:

$$1 + \alpha(L; \theta) = (1 - L)^{-d} \frac{\chi(L)}{\phi(L)} = (1 - L)^{-d} \frac{(1 + \chi_1 L + \cdots + \chi_q L^q)}{(1 - \phi_1 L - \cdots - \phi_p L^p)}, \qquad (9)$$

setting $\theta = (\chi_1, \ldots, \chi_q, \phi_1, \ldots, \phi_p, d)'$ and assuming that the polynomials in the lag operator $L$ (where $x_{t-1} = L x_t$), $\chi(L)$ and $\phi(L)$, have roots strictly outside the unit circle in the complex plane and never common. Remembering that non-negativity constraints on the $\alpha_i(\theta)$ need not be imposed, one can allow for cyclical behaviour in the $\alpha_i(\theta)$; e.g. by the standard re-parameterization of the AR($p$) coefficients, with $p = 2$, as $\phi_1 = 2\tilde{\phi}_1 cos(\tilde{\phi}_2)$, $\phi_2 = -\tilde{\phi}_1^2$ with $0 \leq \tilde{\phi}_1 < 1$ and $0 \leq \tilde{\phi}_2 \leq \pi$.

## 3.2 Limiting distribution theory of the Gaussian estimator

The Gaussian estimator, originally proposed in RZ, is based on Whittle's function (see Whittle (1962)), i.e. the frequency domain approximation of the Gaussian log likelihood, used as a measure of distance between the periodogram of squared observations and the model spectral density $f_{yy}(\lambda; \psi, \sigma^2)$. In particular, setting $\lambda_j = 2\pi j/T$, we consider the following discrete version of Whittle's function

$$Q_T(\sigma^2, \psi) = \frac{1}{T}\sum_{j=1}^{T-1}\left[\ln(\sigma^8 h_y(\lambda_j; \psi)) + \frac{I_{yy}(\lambda_j)}{\sigma^8 h_y(\lambda_j; \psi)}\right], \tag{10}$$

where $I_{yy}(\omega)$ defines the periodogram based on consecutive $T$ observations of the $y_t$:

$$I_{yy}(\omega) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}(y_t - \bar{y})e^{it\omega}\right|^2, \quad -\pi \leq \omega < \pi \tag{11}$$

with $\bar{y} = 1/T\sum_{t=1}^{T} y_t$. We skip the zero frequency in (10) due to mean correction. Concentrating $\sigma^8$ out yields

$$Q_T(\psi) = \ln\left(\hat{\sigma}_T^8(\psi)\right) + \frac{1}{T}\sum_{j=1}^{T-1}\ln\, h_y(\lambda_j; \psi) + 1\,, \tag{12}$$

setting

$$\hat{\sigma}_T^8(\psi) = \frac{1}{T}\sum_{j=1}^{T-1}\frac{I_{yy}(\lambda_j)}{h_y(\lambda_j; \psi)}.$$

Set $\bar{\Psi}$ equal to the compact closure of $\Psi$. The Gaussian estimator $\hat{\psi}_T$ is defined by

$$\hat{\psi}_T = argmin_{\psi \in \bar{\Psi}} Q_T(\psi)\,.$$

The main motivation for employing Whittle's function is that it naturally takes into account the asymptotic behaviour of the ACF of the $y_t$, as the sample size goes to infinity. Therefore we expect it to be very sensitive to the degree of dependence of the process in a second-order sense. Indeed, in a semiparametric framework Robinson (1995a) showed how the most efficient (semiparametric) estimator of the memory parameter is obtained via a local Gaussian log likelihood. Moreover, Whittle's function does not depend on initial conditions and its computation and asymptotic behaviour is independent from the invertibility of the model. On the other hand, the Gaussian estimator requires knowledge of the precise functional form of the model spectral density (cf. Theorem 1). Moreover, independently of the distributional assumption made on the $\epsilon_t$, the $y_t$ can never be Gaussian by construction and thus the Gaussian estimator can never achieve asymptotic efficiency in contrast to MLE.

One cannot implement the Gaussian estimator by employing the levels $x_t$ as the observable. In this case, parameter non-identification arises because the $x_t$ display a perfectly flat power spectrum (cf. Theorem 1). In principle, different instantaneous transformations may be considered but, using the squares $y_t = x_t^2$ as the observable represents the simplest feasible approach.

This estimation approach can be extended considering a multivariate version. Let us take the bivariate process $z_t = (x_t, y_t)$ as the observable. Under the presumption that the $z_t$ are jointly Gaussian, one obtains the 'bivariate' Gaussian estimator by minimizing the 'bivariate' Whittle's function. Using this 'bivariate' Gaussian estimator one achieves a certain efficiency gain with respect to the (univariate) Gaussian estimator (see Zaffaroni (1997)). Within an ARCH-type setting, Meddahi and Renault (1997) show that using a time domain 'bivariate' Gaussian PMLE, based on the correct specification of the first and second conditional moment of the observable, one achieves the generalized method of moment efficiency bound in the sense of Chamberlain (1987). This suggests that a similar result might hold for the 'bivariate' Gaussian estimator. However, in this paper we will focus only on the asymptotics of the (univariate) Gaussian estimator and leave the analysis of the multivariate Gaussian estimator, and the related efficiency issue, to future research.

In order to establish the asymptotic properties of the Gaussian estimator $\hat{\psi}_T$ we introduce the following assumptions. Let $k, K$ denote arbitrary constants, possibly varying in each context. They might depend on $\psi$, so that e.g. $k = k(\psi)$, but for the sake of simplicity we will not make this explicit.

**Assumptions B.** *Let $\psi_0$ be an interior point of $\bar{\Psi}$. For any $\theta, \theta^* \in \Theta$:*

B$_1$ $\quad \alpha_j(\theta) \sim k j^{d(\theta)-1}\,,\ 0 <\mid k \mid < \infty\,,$ *as* $j \to \infty$,

*where the function* $d : \Theta \to (0, \frac{1}{2})$ *is continuous.*

B$_2$ $\quad \mid \alpha_j(\theta) - \alpha_{j+1}(\theta) \mid \leq k \dfrac{\mid \alpha_j(\theta) \mid}{j}\,, \forall j > J$, *some* $J < \infty\,, 0 < k < \infty$.

B$_3$ $\quad \dfrac{\partial \alpha_j(\theta)}{\partial \theta_i} \sim k_i' \, L\,(j) j^{d(\theta)-1}\,, 0 \ <\mid k_i' \mid < \infty\,,\ \ as\ j \to \infty,\ \ for\ i = 1, 2, ..., p.$

B$_4$ $\quad \left| \dfrac{\partial \alpha_j(\theta)}{\partial \theta_i} - \dfrac{\partial \alpha_{j+1}(\theta)}{\partial \theta_i} \right| \leq \dfrac{k'}{j} \left| \dfrac{\partial \alpha_j(\theta)}{\partial \theta_i} \right|,$

$\forall j > J',\ some\ J' < \infty\,, 0 < k' < \infty,\ \ for\ i = 1, 2, ..., p.$

B$_5$ $\quad \dfrac{\partial^2 \alpha_j(\theta)}{\partial \theta_{i_1} \partial \theta_{i_2}} \ \sim\ k''_{i_1,i_2} L_{i_1}(j)\, L_{i_2}(j)\, j^{d(\theta)-1},\ 0 <\mid k''_{i_1,i_2} \mid < \infty,\ \ \ as\ j \to \infty,$

$for\ i_1, i_2 = 1, 2, ..., p.$

B$_6$ $\quad \left| \dfrac{\partial^2 \alpha_j(\theta)}{\partial \theta_{i_1} \partial \theta_{i_2}} - \dfrac{\partial^2 \alpha_{j+1}(\theta)}{\partial \theta_{i_1} \partial \theta_{i_2}} \right| \leq \dfrac{k''}{j} \left| \dfrac{\partial^2 \alpha_j(\theta)}{\partial \theta_{i_1} \partial \theta_{i_2}} \right|,\ \forall j > J''$

$,\ some\ J'' < \infty,\ 0 < k'' < \infty\,,\ \ for\ i_1, i_2 = 1, 2, ..., p.$

B$_7$ $\quad \alpha_j(\theta) \neq \alpha_j(\theta^*), \forall j > J''',\ \ some\ J''' < \infty\ and\ \theta \neq \theta^*.$

B$_8$ $\quad$ *There are* $2(p+1)$ *coefficients* $\alpha_{u_i}(\theta)$ *and* $\partial \alpha_{u_i}(\theta)/\partial \theta\, (i = 1, \ldots, p+1)$ *such that* $\partial \bar{\gamma}_{yy}(u_j; \psi)/\partial \psi\, (j = 1, \ldots, q)$, *with* $q \geq p+1$, *are linearly independent vectors,*

*where* $L(j)$, $L_{i_1}(j)$, $L_{i_2}(j)$, $j = 1, 2, ...$ *are either constant or equal to* $\ln(j)$ *and* $J, J', J'', J'''$ *are equal to some finite integers.*

**Remarks B.** (a) Assumption $B_1$ defines the $\alpha_i(\theta)$ to behave asymptotically as the MA coefficients of stationary ARFIMA$(p, d, q)$, with real $0 < d < 1/2$ and integers $p, q$. The imposing of an exact rate in $B_1$, rather than an upper bound, plays a crucial role in the asymptotic results, e.g. together with $B_7$ it guarantees identification of the model. Assumption $B_2$ implies that the sequence $\{\alpha_i(\theta)\}$ converges to zero smoothly enough or, more formally, that it is quasi monotonically convergent to zero (see Yong (1974, p. 2)). Thus, besides determining the behaviour near the zero frequency of the $f_{yy}(\lambda; \psi, \sigma^2)$, $B_2$ imposes a Lipschitz condition (see Zygmund (1977, p. 42)) on the model spectrum in the interval $(0, \pi]$. Assumptions $B_3$, $B_4$ and $B_5$, $B_6$ impart the same degree of smoothness as $B_1$ and $B_2$ to $\partial f_{yy}(\lambda; \psi, \sigma^2)/\partial \psi$ and to $\partial^2 f_{yy}(\lambda; \psi, \sigma^2)/\partial \psi \partial \psi'$ respectively. Assumptions $B_5$, $B_6$ are needed in order to establish the CLT of the Gaussian estimator, imposing the required degree of smoothness to the Hessian. $B_7$ represents an identification condition. Assumption $B_8$ guarantees that the Hessian is nonsingular and it is a rather mild identification condition, ruling out over-parameterizations. $B_8$ is based on the short run behaviour of the $\alpha_i(\theta)$ and derivative of, in contrast to $B_3$ which implies that the $\partial \alpha_u(\theta)/\partial \theta$ are, as $u \to \infty$ , asymptotically linearly dependent. We are not assuming, for the sake of simplicity, the possibility of slowly varying factors in the definition of the $\alpha_i(\theta)$ but they can nevertheless arise as a by-product (expressed by the logarithm function $\ln(\cdot)$) when differentiating. Finally, note that nowhere are non-negativity

constraints imposed. This implies a great deal of flexibility when considering finite-parameterization the $\alpha_i(\theta)$ compatible with Assumptions $B$.
(b) Time domain regularity assumptions, such as Assumptions $B$, are not common in the long memory parametric (see Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), Hosoya (1997) among others) and semi-nonparametric literature, with the exception of Robinson (1994a). Indeed, regularity conditions are usually expressed in terms of a certain degree of smoothness in the model power spectrum.(Fox and Taqqu (1986, remark at p.529) acknowledge that, in certain cases, Tauberian theorems can be applied to obtain the asymptotic behaviour of the Fourier coefficients of the model power spectrum and its derivatives.) However, whereas in a semi-nonparametric framework those assumptions represent a natural choice, they may not be as well motivated in a parametric framework. In fact, it might be very difficult in general, within a specific nonlinear parameterization, to check that the model power spectra driven by this parameterization satisfy such frequency domain assumptions. In our nonlinear parametric framework the statistical properties of the model are completely determined by the behaviour of the unobservables $\epsilon_t$ and of the coefficients $\alpha_i(\theta)$. Therefore it seems natural to impose regularity conditions as stated by Assumptions $B$ above. These conditions appear very easy to check. We show that the MA coefficients of ARFIMA$(p, d, q)$ do satisfy Assumptions $B$, representing a simple yet feasible parameterization of the model (cf. appendix E).

The asymptotic properties, consistency and asymptotic normality, of the Gaussian estimators $\hat{\psi}$ and $\hat{\sigma}^2_T(\widehat{\psi}_T)$ follow.

**Theorem 2** *Under Assumptions $A, B_1, B_2, B_3, B_4, B_7$, as $T \to \infty$,*

$$\begin{aligned} \hat{\psi}_T &\to_{a.s.} \psi_0, \\ \hat{\sigma}^2_T(\widehat{\psi}_T) &\to_{a.s.} \sigma^2, \end{aligned}$$

*where $\to_{a.s.}$ denotes convergence almost surely (a.s.).*

**Proof:** see appendix B.2.

We need to reinforce Assumption $A$ as follows.

**Assumption A'.** *The $\epsilon_t$ satisfy Assumption $A$ with $E\epsilon_t^8 < \infty$ and zero cumulants of order $s = 5, ..., 8$.*

**Theorem 3** *Under Assumptions $A', B$, as $T \to \infty$,*

$$T^{\frac{1}{2}}(\hat{\psi}_T - \psi_0) \to_d \mathcal{N}_{p+1}\left(0, M^{-1}VM^{-1}\right),$$

*where*
$$M(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} N(\lambda;\psi)N(\lambda;\psi)'d\lambda,$$
*with*
$$N(\lambda;\psi) = \frac{\partial \ln\, h_y(\lambda;\psi)}{\partial\psi} - \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial \ln\, h_y(\omega;\psi)}{\partial\psi}d\omega\right],$$
*and*
$$\begin{aligned} V(\psi) \quad &= \frac{1}{\pi}\int_{-\pi}^{\pi} g(\omega;\psi)g(\omega;\psi)'h_y^2(\omega)d\omega \\ &+\frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1;\psi)g(\omega_2;\psi)'\overline{Q}_{yyy}(-\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2\,, \end{aligned}$$

*with* $g(\lambda;\psi) = \frac{\partial}{\partial\psi}h_y^{-1}(\lambda;\psi)$*,* $M = M(\psi_0)$*,* $V = V(\psi_0)$ *and* $\overline{Q}_{yyy}(.,.,.) = Q_{yyy}(.,.,.)/\sigma^{16}$*,* $Q_{yyy}(.,.,.)$ *denoting the trispectrum of the* $y_t$ *(the Fourier transform of the fourth-order cumulants of the* $y_t$*),* $\rightarrow_d$ *denotes convergence in distribution and* $\mathcal{N}_q(\cdot,\cdot)$ *a q-dimensional normal r.v.*

**Proof;** see appendix B.3.
**Remarks 2 and 3.**(a) The important aspect of Assumption $A'$ is that it imposes the strong *i.i.d.* property on the $\epsilon_t$, imparting strict stationarity and ergodicity to the $y_t$. Zero restrictions on cumulants are made for simplicity's sake but play no substantial role in the proofs and the results of Theorem 2 and Theorem 3 are unchanged with constant cumulants and a bounded eighth moment.
(b) A strong law of large number (LLN) for $\hat{\psi}_T$ and $\hat{\sigma}_T^2(\widehat{\psi_T})$ is obtained adapting the Hannan (1973) approach although, unlike the scalar-valued Gaussian and linear long memory case (see e.g. Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990)), we cannot employ Hannan (1973) directly. In fact, as a by-product of the nonlinearity, the model features a non-standard parameterization in the sense that $f_{yy}(\lambda;\psi,\sigma^2)$ does not seem to be factored, unlike the linearly regular process case (see Hannan (1970, Chapter II-5)). In fact, although the Wold representation theorem implies that $f_{yy}(\lambda)$ can be written, under Assumption $A$, as

$$f_{yy}(\lambda) = \frac{\tau}{2\pi}\mid 1 + \sum_{j=1}^{\infty}\beta_j e^{ij\lambda}\mid^2, \quad -\pi \leq \lambda < \pi, \tag{13}$$

the determination of the $\beta_j$ and $\tau$ as closed-form, functionally independent, functions of the $\alpha_i$ and $\sigma^2, \kappa, \rho^2, \mu$, appears very difficult, practically unfeasible. Robinson (1978) discusses several cases where the spectral density of the observable is not easily factored and provides the asymptotics of the Gaussian estimator for those cases without assuming linearity. However, only

short memory specifications of the nonlinear MA might be covered in that framework as strong smoothness assumptions on the model spectrum are imposed, ruling out long memory.
(c) The theoretical result, on which Theorem 3 is based, is a CLT for quadratic forms in the $y_t$, a non-Gaussian and nonlinear sequence with long memory, which the statistical literature on Gaussian estimation for linear long memory processes does not cover. Limit laws for quadratic form in scalar-valued long memory Gaussian and linear processes are in Fox and Taqqu (1987) and Giraitis and Surgailis (1990), respectively. The former uses Gaussianity in an essential way, employing the exact expression for the cumulants of a quadratic form in Gaussian variates (see Grenander and Szegö (1958)) whereas the latter, though relaxing the Gaussianity assumption, exploits the simple structure of the spectrum and trispectrum of a linear process (cf. (13)). More recently Heyde and Gay (1993) and Hosoya (1997) have developed CLT for vector-valued linear non-Gaussian processes under weaker conditions on the innovations, the latter also allowing for non-standard conditions and non-factorization in the model spectral density. However, those results are ultimately based on some martingale approximation, making a basic use of the maintained structure of linearity and imposing some form of mixing condition on the linear innovations, stronger than the white-noise property implied by the Wold representation. Our method of proof, instead, takes explicitly into account the nonlinear structure of the model and consists in approximating the quadratic form in the $y_t$ (embedded in Whittle's function) with another quadratic form, asymptotically equivalent to the former, for which a CLT follows using standard results. In order to establish this approximation we use certain results of the asymptotic behaviour of the trace of Toeplitz matrices (see Fox and Taqqu (1987)).
(d) The relevance of Theorem 3 stems from two reasons. First, the long memory feature of the squares $y_t$ might have generated 'per se' non-standard asymptotic distribution such as nonlinear functionals of fractional Brownian motion (see Taqqu (1975)), or a rate of convergence depending on the degree of memory of the process, as for the case of the sample mean of $T$ consecutive observations of a long memory process. Instead, using the Gaussian estimator standard asymptotic inference can be made on the nonlinear MA such as Wald-type tests. This is due to the typical compensation effect of Whittle's function which smoothes the possible lack of square integrability of $f_{yy}(\lambda)$. Second, and more importantly, only for a limited number of volatility models the asymptotic properties of certain estimators (usually the Gaussian PMLE) have been established (see Weiss (1986) for ARCH($p$) and Lee and Hansen (1994), Lumsdaine (1996) GARCH$(1,1)$). Long memory parameterizations are always excluded. This lack of results in fact characterizes all long memory volatility models, including SV models. The only exception is

Breidt, Crato, and deLima (1998) who establish consistency of the Gaussian estimator for the long memory SV model of Harvey (1998), exploiting the linearity of ln $y_t$ for long memory SV $x_t$.
(e) As a simple case of Robinson (2001), (7) generalizes to

$$\mathrm{cov}\,(g(\epsilon_t\, h_{t-1}), g(\epsilon_{t+u}\, h_{t+u-1})) \sim k\, u^{2d-1} \quad \text{as } u \to \infty,$$

for Gaussian $\epsilon_t$ and any arbitrary function $g(\cdot)$ satisfying $Eg^2(\epsilon_t\, h_{t-1}) < \infty$ and $E\, h_{t-1}\, g(\epsilon_t\, h_{t-1}) \neq 0$, with $g(x\, y)$ either equal to $g(x)g(y)$ or $g(x) + g(y)$. Particular cases are $g(t) = \mid t \mid^{\alpha}$, $\alpha > 0$ and $g(t) = \ln t^2$. Therefore, for nonlinear MA the $y_t$, the $\mid x_t \mid$ and the $\ln y_t$ have precisely the same degree of long memory. As $\ln y_t$ is defined over the entire real line and thus closer to a Gaussian variate than $y_t$, it would probably be more efficient to estimate the memory parameter $d$ by fitting the spectral density of the $\ln y_t$ rather than $f_{yy}(\lambda)$. Moreover, Wright (2000) indicates that, when the levels $x_t$ are conditionally leptokurtic, a non-negligible small sample negative bias arises systematically when estimating semiparametrically the memory parameter $d$ using the $y_t$ as the observables, compared with the case when the $|\, x_t\, |$ or the $\ln y_t$ are used. Although in principle other instantaneous transformations of nonlinear MA $x_t$ could be used besides $y_t = x_t^2$, the Gaussian estimator requires knowledge of the precise functional form of the model spectral density for the transformed data. This task appears cumbersome, if not practically unfeasible, to be derived for the $|\, x_t\, |$ and the $\ln y_t$ of nonlinear MA.

### 3.3 Testing long memory dynamic heteroskedasticity

When limiting the attention to short memory specification of the model, any LM test for ARCH effects that is asymptotically equivalent to Engle (1982) test provides an efficient way to detect short memory conditional heteroskedasticity. In this section we introduce a family of LM tests specifically designed to have good power against long memory dynamic conditional heteroskedastic alternatives. The test statistics introduced here are developed assuming that under the alternative hypothesis the observable is described by the nonlinear MA. Nonparametric (see e.g. Lo (1991)) and semiparametric approaches (see e.g. Robinson (1994b,1995a,1995b)) can still be applied but with the usual low performance in terms of power. In a parametric framework the LM tests of Robinson (1991) can be considered but they are derived under ARCH-type alternatives. The test we propose follows the approach of Robinson (1991) but, by construction, it has more local asymptotic power when the alternative is represented by the nonlinear MA. This is, obviously, particularly important when fitting the nonlinear MA to the data. The following assumptions concerning both the $\alpha_i(\theta)$ and the $\epsilon_t$ are needed.

**Assumptions C.**

$C_1$ *The coefficients* $\{\alpha_i(\theta)\}$ *are invertible functions of a* $p \times 1$ *vector* $\theta$ *such that* $\alpha_i(\theta) = 0\,(i \geq 1)$ *if and only if* $\theta = 0$. *The partial derivatives of* $\alpha_i(\theta)$ *are square summable for* $\theta = 0$ *where we set* $\tau_i(\theta) = \frac{\partial}{\partial\theta}\alpha_i(\theta)$ *with* $\tau_i = \tau_i(0)$.

$C_2$ *The matrix* $\Gamma = \sum_{i=1}^{\infty} \tau_i \tau_i'$ *is invertible.*

$C_3$ $\rho^2 = 1$.

**Assumption D.** *The process* $\{\epsilon_t\}$ *behaves as a Gaussian noise sequence up to the eighth moment.*

We define the null hypothesis as

$$H_0: \quad \alpha_i(0) = 0, \;\; i \geq 1,$$

so that under $H_0$ both the $x_t$ and the $y_t$ are not autocorrelated.

Assumption $C_1$ guarantees that the null hypothesis can be equivalently characterized in terms of the $\alpha_i(\theta)$ or, alternatively, in terms of $\theta$. Hence we can write $H_0: \;\; \theta = 0$. Assumptions $C_2$ and $C_3$ are minimal identification conditions. In fact, $x_t = \mu + \rho\epsilon_t$ under $H_0$ implying $\mathrm{var}(x_t) = \rho^2\sigma^2$ and $\rho$ is not identified under $H_0$. Finally, note that nowhere in Assumption $D$ did we assume the $\epsilon_t$ to be an *i.i.d.* sequence.

Departures from $H_0$ do not necessarily display long memory, but express behaviours for the $\alpha_i(\theta)$ with at least a finite number of them non-zero, including any short memory parameterization. Note that any sequence $\alpha_i(\theta)$ satisfying Assumption $C_1$, $C_2$ is compatible with both $H_0$ and Assumptions $B$.

Setting $w = \mu^2/\sigma^2$, the parameterized model is described by the parameter vector $(w, \sigma^2, \theta')'$. The LM test statistic for testing $H_0$ is (cf. appendix D):

$$\Phi = \phi'\phi, \tag{14}$$

with

$$\phi = \phi_{T-1}(\hat{\sigma}^2, \hat{w}) = (T\, e(\hat{w})\, \hat{\sigma}^2\, \Gamma_{T-1})^{-\frac{1}{2}} \sum_{i=1}^{T-1} \tau_i\, \eta_i(\hat{\sigma}^2, \hat{w}), \tag{15}$$

setting

$$\eta_i(\sigma^2, w) = \sum_{t=1+i}^{T} (x_{t-i} - \bar{x})\, X_t(\sigma^2, w),$$
$$X_t(\sigma^2, w) = \left( (\frac{y_t}{\sigma^2} - w - 1) \left( (\frac{y_t}{\sigma^2} - w - 1) \frac{(1+w)}{(1+2w)} + 1 \right) - 2(1+w) \right),$$

and $e(w) = 9+(65+150w^3+338w^2+280w)/(1+2w)^2$, $\Gamma_T = \sum_{i=1}^T \tau_i\tau_i'$ where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (x_t - \bar{x})^2$, $\bar{x} = \frac{1}{T}\sum_{i=1}^T x_i$, $\hat{w} = \bar{x}^2/\hat{\sigma}^2$. One can use $\Gamma$ instead of $\Gamma_T$ if the former has a closed-form expression. As a discrete convolution is involved, we can make use of the Fast Fourier Transform (FFT) with the usual computational gain. Indeed, a frequency domain version of the LM test can be computed:

$$\tilde{\Phi} = \tilde{\phi}'\tilde{\phi}\,, \tag{16}$$

with

$$\tilde{\phi} = (T\, e(\hat{w})\hat{\sigma}^2\Gamma_T)^{-\frac{1}{2}}\frac{1}{T}\sum_{a=0}^{T-1,\,*}\tilde{\tau}(\omega_a)\tilde{x}(\omega_a)\tilde{X}(-\omega_a)\,,$$

setting $\tilde{X}(\omega) = \sum_{a=0}^{T-1} X_{a+1}(\hat{\sigma^2}, \hat{w})e^{ia\omega}$, $\tilde{x}(\omega) = \sum_{a=0}^{T-1}(x_{a+1} - \bar{x})e^{ia\omega}$, $\tilde{\tau}(\omega) = \sum_{a=0}^{\infty}\tau_{a+1}e^{ia\omega}$, with $\sum^*$ meaning that the summation excludes values of $\omega_a$ for which $\tilde{\tau}(\omega_a)$ is unbounded. The matrix $\Gamma$ can now be expressed as $\Gamma = \frac{1}{4\pi}\int_{-\pi}^{\pi}\tilde{\tau}(\omega)\tilde{\tau}'(-\omega)d\omega$, which, in turn, can be approximated by its Riemann sum. Thus

**Theorem 4** *Under the null hypothesis $H_0$ and Assumptions C and D,*

$$\Phi \to_d \chi_p^2 \quad as \quad T \to \infty,$$

*where $\chi_p^2$ denotes a central chi-square r.v. with $p$ degrees of freedom.*

**Proof:** see appendix D.

**Remarks 4.** (a) A standard asymptotic distribution is obtained though the test has been designed for possibly long memory alternatives.

(b) The two test statistics $\Phi$ and $\tilde{\Phi}$ are asymptotically equivalent, given that $\Phi - \tilde{\Phi} = o_p(1)$, so Theorem 4 applies to $\tilde{\Phi}$ as well.

(c) From a computational point of view, calculating $\Phi$ involves no more than a few summations and the inversion of a matrix of dimension $p \times p$ ($\theta$ is of dimension $p \times 1$) and thus not a function of $T$. The LM ARCH test of Engle (1982), instead, requires inverting a matrix of dimension $r \times r$, with $r$ being the number of lags considered. It follows that taking into account the behaviour at long lags, for example in order to capture long memory behaviour, makes the LM ARCH test unfeasible.

(d) The statistic $\Phi$ is based on the Gaussian log likelihood, so it is in fact a Gaussian pseudo LM test. Unlike the estimation part, the time domain Gaussian log likelihood is used as the objective function. In fact, under $H_0$ the model is trivially invertible given that $x_t = \mu + \epsilon_t$.

(e) Theorem 4 effectively establishes that $\phi$ is converging in distribution to a multivariate standard normal so that, when $p = 1$, a one-sided test of $H_0$ against the alternative (say) $\theta > 0$ can be performed with the usual gain in power (cf. Robinson (1991)).

(f) The obvious, mild condition for consistency of $\Phi$ is $\sum_{i=1}^{\infty} \tau_i E(\eta_i) \neq 0$ when $H_0$ is false.

In terms of the assumptions made, $D$ imposes a relatively stringent structure on the moments of the unobservable $\epsilon_t$. For this reason we propose a robustified LM test statistic, which has the same asymptotic distribution under $H_0$ as the $\Phi$ statistic, but allowing a great deal of heterogeneity in the $\epsilon_t$. Details are reported in appendix D.1.

# 4 ARMA with nonlinear MA innovations

To relax the hypothesis of conditionally constant first moment of the levels $x_t$ we extend (1) to

$$x_t = \sum_{j=0}^{\infty} \beta_j u_{t-j}, \ \ \beta_0 = 1, \ \ \sum_{j=0}^{\infty} \beta_j^2 < \infty, \ \ t \in \mathbb{Z}, \tag{17}$$

with nonlinear MA innovations $u_t$ described by (2). With no loss of generality, $E(x_t) = 0$ under (17) and Assumption $A$.

For parametric estimation of the $\beta_j$, a flexible and computationally convenient parameterization is provided by the ARMA$(r, q)$ model, whereby

$$x_t = \omega_{1,0} x_{t-1} + \omega_{2,0} x_{t-2} + ... + \omega_{r,0} x_{t-r} + u_t + \xi_{1,0} u_{t-1} + .. + \xi_{q,0} u_{t-q}, \tag{18}$$

yielding $\beta_j = \beta_j(\zeta_0)$ with $\zeta_0 = (\omega_0', \xi_0')'$ for constant coefficients

$\omega_0 = (\omega_{1,0}, ..., \omega_{r,0})'$, $\xi_0 = (\xi_{1,0}, ..., \xi_{q,0})'$, such that the polynomials $\Omega_0(z) = 1 - \omega_{1,0}\, z - ...\omega_{r,0}\, z^r$ and $\Xi_0(z) = 1 + \xi_{1,0} z + ...\xi_{q,0} z^q$ satisfy

$$\Xi_0(z)\Omega_0(z) \neq 0, \ \ |\, z \,| \leq 1, \tag{19}$$

implying (17). Clearly $\beta(L; \zeta_0) = \sum_{i=0}^{\infty} \beta_i(\zeta_0) L^i = \Xi_0(L)/\Omega_0(L)$.

Model (18) with (2) could be estimated by a two-stage approach. In the first stage, one estimates the conditional mean ARMA parameters $\zeta_0$. Given the martingale difference property of the innovations $u_t$ and $\gamma_{uu}(0) < \infty$, implied by Assumption $A$, the observable $x_t$ have spectral density now equal to

$$f_{xx}(\lambda) = \frac{\gamma_{uu}(0)}{2\pi} \frac{|\ \Xi_0(e^{i\lambda})\ |^2}{|\ \Omega_0(e^{i\lambda})\ |^2} = \frac{\gamma_{uu}(0)}{2\pi} p_x(\lambda; \zeta_0), \ \ -\pi \leq \lambda < \pi. \tag{20}$$

Given (20) the mean parameters could be estimated by the Whittle estimator $\hat{\zeta}$, defined by

$$\hat{\zeta}_T = argmin_{\zeta \in \bar{\mathcal{Z}}}\, P_T(\zeta), \tag{21}$$

setting

$$P_T(\zeta) = \frac{2\pi}{T} \sum_{j=1}^{T-1} \frac{I_{xx}(\lambda_j)}{p_x(\lambda_j;\zeta)},$$

where $p_x(\lambda;\zeta) =\mid \beta(e^{i\lambda};\zeta) \mid^2$ defines the parametrized model spectrum for the $x_t$ (up to a constant of proportionality), $\mathcal{Z} = \{\zeta \in \mathbb{R}^{r+q} \mid \Omega(z)\Xi(z) \neq 0$ for $\mid z \mid \leq 1$ and $\Omega(z)$, $\Xi(z)$ have no common roots $\}$, $\bar{\mathcal{Z}}$ defines the compact closure of $\mathcal{Z}$, and

$$I_{ab}(\lambda) = \frac{1}{2\pi T} \sum_{t,s=1}^{T} (a_t - \bar{a})(b_s - \bar{b}) e^{-i\,(t-s)\,\lambda}, \ \ -\pi \leq \lambda < \pi$$

for any sequences $\{a_t\}$, $\{b_t\}$. Likewise, we define the cross-spectrum (when it exists)

$$f_{ab}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \mathrm{cov}(a_t, b_{t+u}) e^{i\lambda u}, \quad -\pi \leq \lambda < \pi.$$

The asymptotic properties of $\hat{\zeta}_T$ follows.

**Theorem 5** *Under Assumption A, for $\zeta_0$ being an interior point of $\bar{\mathcal{Z}}$, as $T \to \infty$,*

$$\hat{\zeta}_T \to_{a.s.} \zeta_0,$$

*and*

$$T^{\frac{1}{2}}(\hat{\zeta}_T - \zeta_0) \to_d \mathcal{N}_{r+q}\left(0, F^{-1} G F^{-1}\right),$$

*where*

$$F(\zeta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} b(\lambda;\zeta)\, b(\lambda;\zeta)' d\lambda, \tag{22}$$

$$G(\zeta) = 2F(\zeta) + \frac{1}{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} b(\omega_1;\zeta) b(\omega_2;\zeta)' \frac{f_{u^2u^2}(\omega_1+\omega_2)}{\gamma_{uu}^2(0)} d\omega_1 d\omega_2 \tag{23}$$

*and*

$$b(\lambda;\zeta) = \frac{\partial}{\partial \zeta} \ln\, p_x(\lambda;\zeta) \tag{24}$$

*with $F = F(\zeta_0)$, $G = G(\zeta_0)$, $b(\lambda) = b(\lambda;\zeta_0)$ and $f_{u^2u^2}(\lambda)$ defines the power spectrum of the $u_t^2$.*

**Proof:** see appendix C.

**Remarks 5.** (a) Expression (23) gives the general form of the matrix $G(\zeta)$ whenever one considers Whittle estimation of ARMA$(r,q)$ allowing for conditionally heteroskedastic innovations $u_t$ with bounded fourth moment. This includes not only nonlinear MA but also, for example GARCH (imposing bounded fourth moment). (23) simplifies to $G(\zeta) = 2F(\zeta)$ when the $u_t$ are

independent or, more generally, conditionally homoskedastic.
(b) The expression for the power spectrum of the $u_t^2$, $f_{u^2u^2}(\lambda)$ enters into the definition of $G$. The exact expression for $f_{u^2u^2}(\lambda)$ is given in Theorem 1 (setting $\mu = 0$ in $f_{yy}(\lambda)$). However, note that a consistent estimate of $G$ does not require the knowledge of this expression and can be obtained replacing $f_{u^2u^2}(\lambda)$ with $I_{u^2u^2}(\lambda)$ and adapting Taniguchi (1982).
(c) $P_T(\hat{\zeta}_T)$ represents a strongly consistent estimator of $\gamma_{uu}(0)$ but, although $\gamma_{uu}(0) = \sigma^4(\bar{\rho}^2 + \delta_{\alpha\alpha}(0))$, this cannot be used to identify and thus estimate the nonlinear MA parameters, as previously noted.
(d) When assuming $q = 0$, for AR($p$) $x_t$, the Whittle estimator of $\zeta_0 = \omega_0$ becomes the Yule-Walker estimator, which has a closed-form expression and is asymptotically equivalent to the least squares estimator (see Brockwell and Davis (1987, section 8.1)).

Given (19) the AR($\infty$) representation of the $x_t$ follows:

$$x_t - \sum_{j=1}^{\infty} \pi_j(\zeta_0) x_{t-j} = u_t,$$

with $1 - \sum_{j=1}^{\infty} \pi_j(\zeta_0) L^j = \Omega_0(L)/\Xi_0(L)$, from which one gets the estimated residuals as

$$\hat{u}_t = x_t - \sum_{j=1}^{t-1} \pi(\hat{\zeta}_T) x_{t-j}, \;\; t = 1, ..., T,$$

where $\sum_{i=a}^{b} c_i = 0$ for $a > b$ for any sequence $\{c_i\}$. Other methods of estimation of the residuals can be used such as the Kalman filter (see Brockwell and Davis (1987, section 12.2)).

Given the $\hat{u}_t$ one then estimates the nonlinear MA parameters $\psi = (\bar{\kappa}, \theta')'$ with the Gaussian estimator, replacing the $y_t$ with the $\hat{u}_t^2$ in section 3. Let

$$\tilde{Q}_T(\psi) = \ln\left(\tilde{\sigma}_T^8(\psi)\right) + \frac{1}{T}\sum_{j=1}^{T-1} \ln\, h_y(\lambda_j; \psi) + 1\,, \tag{25}$$

setting

$$\tilde{\sigma}_T^8(\psi) = \frac{1}{T}\sum_{j=1}^{T-1} \frac{I_{\hat{u}^2\hat{u}^2}(\lambda_j)}{h_y(\lambda_j;\psi)}.$$

The Gaussian estimator $\tilde{\psi}_T$, based on the estimated residuals $\hat{u}_t$, is defined by

$$\tilde{\psi}_T = argmin_{\psi \in \bar{\Psi}} \tilde{Q}_T(\psi)\,.$$

It turns out that $\tilde{\psi}_T$ is still (strongly) $\sqrt{T}$-consistent and asymptotically normally distributed, although with a different, larger asymptotic variance-covariance matrix from the one of $\hat{\psi}_T$ (stated in Theorem 3).

**Theorem 6** *Under the same assumptions made in Theorem 3, as $T \to \infty$,*

$$\tilde{\psi}_T \to_{a.s.} \psi_0,$$

*and*

$$T^{\frac{1}{2}}(\tilde{\psi}_T - \psi_0) \to_d \mathcal{N}_{p+1}\left(0, M^{-1}UM^{-1}\right),$$

*where*

$$U(\psi) = V(\psi) + H(\psi)F^{-1}GF^{-1}H(\psi)', \quad U = U(\psi_0),$$

*setting*

$$H(\psi) = \frac{2}{\pi\sigma^8}\int_{-\pi}^{\pi} g(\lambda;\psi)\,\mathcal{R}e(f_{u^2\,uv'}(\lambda))d\lambda$$

*with the $(r+q)$-vector valued random variables $v_t = \sum_{j=1}^{\infty}\vartheta_j x_{t-j}$, $\vartheta_j(\zeta) = \partial\pi_j(\zeta)/\partial\zeta$, $\vartheta_j = \vartheta_j(\zeta_0)\,(j \geq 1)$, and with $M, V$ and $g(\lambda;\psi)$ defined in Theorem 3 and $F, G$ defined in Theorem 5.*

**Proof:** see appendix C.
**Remarks 6.** (a) The shape of the asymptotic variance-covariance matrix of $\tilde{\psi}_T$ shows that there is an efficiency loss when estimating the residuals compared with the case when the levels $x_t$ are martingale difference themselves.
(b) A more efficient estimation method is based on joint estimation of $\zeta_0$ and $\psi_0$, such as by using the bivariate Gaussian estimation approach mentioned in section 3.2. The two-stage procedure here proposed, though, appears very appealing computationally compared with practical use of the bivariate Gaussian estimator. Moreover, as discussed above, full efficiency can never be attained by the bivariate Gaussian estimator either. Finally, having established the asymptotics for the univariate Gaussian estimator, the incremental mathematical achievement for the bivariate Gaussian estimator will not be significant.
(c) The result can be easily extended to the case of exogenous regressors such as $x_t = \beta' z_t + \beta(L,\zeta_0)u_t$ under suitable regularity conditions on the $z_t$ (see Hannan (1973, section 4)).
(d) Using $\hat{\zeta}_T$ and $\tilde{\psi}_T$ a consistent estimate of $U$ can be obtained, replacing $f_{u^2\,uv'}(\lambda)$ with $I_{u^2\,u\hat{v}'}(\lambda)$, setting $\hat{v}_t = \sum_{j=1}^{t-1}\vartheta_j(\hat{\zeta}_T)x_{t-j}$, and adapting Taniguchi (1982).

# 5 Montecarlo experiments

We have performed a relatively extensive Montecarlo exercise to assess the small sample performance of the Gaussian estimator and of the LM test. We have considered both the case of martingale difference and ARMA observables. In the latter case, we have also reported the Montecarlo results of the first-stage estimation of ARMA parameters. We consider the following exact, up to the second moment, simulation approach. First, we simulate $T+1$ innovations $\epsilon_t \, (0 \leq t \leq T)$ as *i.i.d.* draw from a Student's $t$ distribution with $n$ degrees of freedom, variance 1 and fourth-order cumulant $\bar{\kappa}$. Normally distributed $\epsilon_t$ are obtained when $n = \infty$ $(\bar{\kappa} = 0)$. Recall that $\theta = (\phi', \chi', d)'$ and $\zeta = (\omega', \xi')'$. $T$ defines sample size. Then we construct the sequence $u_t \, (1 \leq t \leq T)$ as

$$u_t = \epsilon_t \left( \sigma \bar{\rho} + \sqrt{\delta_t(\theta)} \epsilon_0 + \sum_{j=1}^{t-1} \alpha_j(\theta) \epsilon_{t-j} \right), \quad 1 \leq t \leq T,$$

setting

$$\delta_t(\theta) = \begin{cases} \sum_{j=1}^{\infty} \alpha_j^2(\theta), & t = 1, \\ \delta_{t-1}(\theta) - \alpha_{t-1}^2(\theta), & t \geq 2. \end{cases}$$

When considering the simple case of martingale difference levels $(\zeta = 0)$ then set $x_t = u_t \, (1 \leq t \leq T)$ otherwise

$$x_t = \sqrt{\eta_t(\zeta)} u_1 + \sum_{j=0}^{t-2} \beta_j(\zeta) u_{t-j}, \quad 1 \leq t \leq T,$$

setting

$$\eta_t(\zeta) = \begin{cases} \sum_{j=0}^{\infty} \beta_j^2(\zeta), & t = 1, \\ \eta_{t-1}(\zeta) - \beta_{t-2}^2(\zeta), & t \geq 2. \end{cases}$$

Throughout the exercise we always consider 1000 replications and sample size $T$ equal to 1024, 2048, 4096.

At first we assess the performance of the Gaussian estimator assuming martingale difference levels $x_t = u_t$. We focus on the following parameterization:

$$\alpha(L; \theta) = \frac{(1-L)^{-d}}{(1 - \phi_1 L)}, \;\; 0 < d < 1/2, \;\; |\phi_1| < 1, \tag{26}$$

yielding $\psi = (\bar{\kappa}, \phi_1, d)$, with $\bar{\kappa} \in \{0,\, 1.5,\, 3.0\}$ ($n$ equal to infinity, 8 and 6 respectively), $\phi_1 \in \{-0.8,\, -0.4,\, 0.4,\, 0.8\}$ and $d \in \{0.15,\, 0.25,\, 0.45\}$. We set $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$. Tables 1.1, 1.2 and 1.3 reports the sample mean and its standard deviation of the estimates across the 1000 replications.

The results show that for all parameters increasing sample size improves remarkably the results in terms of bias reduction and accuracy. There is, though, a marked difference in the accuracy of the estimates of $\bar{\kappa}$ compared with the other parameters $\phi_1$, $d$. This is probably a consequence of the fact that the Gaussian estimator is not a likelihood-method and, therefore, information on the conditional distribution of the $x_t$ (unconditional distribution of the $\epsilon_t$) enters only indirectly into Whittle's function, through $\bar{\kappa}$. This contrasts with likelihood-methods, such as PMLE, which by construction take directly into account the conditional distribution of the $x_t$. Estimates of $d$ frequently exhibit a negative bias increasing in absolute value with $\bar{\kappa}$. This agrees with the finding of Wright (2000) in a semiparametric setting (cf. remark (e) to Theorem 3). Although estimation of $\bar{\kappa}$ should potentially reduce this problem, the simulations show often that a significant bias arises. Moreover, estimates of $d$ exhibit a greater accuracy compared with estimates of $\phi_1$. However, the Gaussian estimator shows a good performance with respect to the well-known difficulty of disentangling the effect of the long memory parameter from pure autoregressive parameters in estimation of parametric long memory models. Tables 3.1 and 3.2 report the empirical frequencies for confidence intervals at 95%, based on estimates of the parameter $d$ and their asymptotic standard errors, for parameterization (26). The results show a sizeble improvement, as the sample size increases, for medium and large values of $d$. In contrast, no substancial improvement is obtained for small values of $d$. The results tend to worsen for large values of $\bar{\kappa}$, especially for $d$ greater than 0.15.

Next, we consider the case of AR(1) levels, $x_t = \omega_1 x_{t-1} + u_t$ with $|\omega_1| < 1$, and the nonlinear MA coefficients still given by (26), following the approach indicated in section 4. Thus $\zeta = \omega_1$ and $\psi = (\bar{\kappa}, \phi_1, d)$, where $\omega_1 = 0.5$, $\psi_1 \in \{-0.4, 0.4\}$ and $d \in \{0.15, 0.25, 0.45\}$. The results are presented in Tables 2.1, 2.1 and 2.3. The conditional mean parameter $\omega_1$ appears much more precisely estimated compared with the conditional variance parameters $\psi$. The efficiency loss emerging when estimating $\psi$ in the second stage, based on estimated residuals, does not appear too serious in most cases. Again, estimates of $\bar{\kappa}$ are more imprecise compared with other parameter' estimates. For all parameters bias reduction and accuracy improve as $T$ increases.

Finally, Table 4 presents evidence of the small-sample performance, both in terms of size and power, of the LM test $\Phi$ of section 3.3. The data are simulated as described above, with Gaussian $\epsilon_t$, and

$$\alpha(L;\theta) = (1-L)^{-d}, \;\; 0 < d < 1/2.$$

This yields $\tau_i = 1/i$ (cf. Assumption $C$) and $\Gamma = \pi^2/6$ (cf. Robinson (1991)). Good power arises for moderate and large values of $d$ but it is poor when $d = 0.1$. Moreover, the power decreases when $d$ is very close to 1/2, in the proximity of the non-stationary region, possibly the effect of imposing the boundary condition. Overall, the results improve markedly as $T$ increases.

# 6 Empirical application

In this section we assess the performance of the nonlinear MA to fit time series of observed asset returns using the Gaussian estimator. The results are compared with the performance of its most natural competitor, the ARCH($\infty$)

of Robinson (1991):

$$x_t = \nu + \epsilon_t \sigma_t, \;\; t \in \mathbb{Z}, \tag{27}$$

$$\sigma_t^2 = \tau + \sum_{j=1}^{\infty} \psi_j y_{t-j}, \;\; a.s., \;\; \sum_{j=1}^{\infty} \psi_j < \infty, \tag{28}$$

where $\tau$, $\psi_j \geq 0 \, (j \geq 1)$ and the $\epsilon_t$ satisfy Assumption $A$.

Various stationarity conditions and memory properties of ARCH($\infty$) are discussed in RZ, Giraitis, Kokoska, and Leipus (1998) and Zaffaroni (2000). In particular, long memory is ruled out when $Ex_t^2 < \infty$. Imposing $Ey_t^2 < \infty$, the spectral density for ARCH($\infty$) $y_t$ can be factored exhibiting a representation (13), in contrast to nonlinear MA. The (necessary and sufficient) condition for covariance stationarity of ARCH($\infty$) $y_t$ is (using the notation of section 2):

$$0 \leq \Sigma = \frac{2\sigma^4 \tau^2}{\Psi^2(1)} \left( 1 - 2\sigma^4 \sum_{u=-\infty}^{\infty} \delta_{\psi\psi}(u) (\delta_{\gamma\gamma}(u) + \gamma_u) \right)^{-1} < \infty, \tag{29}$$

setting $\Psi(L) = 1 - \sigma^2 \sum_{i=1}^{\infty} \psi_i L^i$, $\Gamma(L) = 1 + \sum_{i=1}^{\infty} \gamma_i L^i = \Psi(L)^{-1}$ and $\kappa = 0$ (see Zaffaroni (2000, Theorem 4)). When (29) holds, ARCH($\infty$) $y_t$ have a well-defined power spectrum:

$$f_{yy}(\lambda) = \frac{\Sigma}{2\pi} \, | \, \Psi(e^{i\lambda}) \, |^{-2}, \;\; -\pi \leq \lambda < \pi. \tag{30}$$

By finite-parameterizing the $\psi_i$, inference on ARCH($\infty$) can be performed using the Gaussian estimator of section 3.2. Giraitis and Robinson (1998) establish consistency and asymptotic normality of the Gaussian estimator of ARCH($\infty$). Given our emphasis on long memory parameterizations, we consider the following parsimonious specification for the $\psi_i$:

$$1 - \sum_{j=1}^{\infty} \psi_j(\theta) L^j = (1-L)^d \frac{\phi(L)}{\chi(L)} = (1-L)^d \frac{(1 - \phi_1 L - . - \phi_p L^p)}{(1 + \chi_1 L + . + \chi_q L^q)}, \tag{31}$$

where $\theta = (\chi_1, ., \chi_q, \phi_1, ., \phi_p, d)'$, with $d > 0$ and, hereafter, we assume that $\phi(L)$, $\chi(L)$ have all zeros outside the unit circle on the complex plane. (These stationarity and invertibility conditions can be easily imposed using the mapping between the ARMA coefficients and the partial autocorrelation function (see Monahan (1984)).) Subsequently, one gets $\Sigma(\theta)$ from (29) replacing the $\psi_i$ with the $\psi_i(\theta)$ and

$$\Psi(L; \theta) = 1 - \sigma^2 \sum_{j=1}^{\infty} \psi_j(\theta) L^j. \tag{32}$$

Condition (29) and parameterization (31) require $\sigma^2 < 1/2$ (cf. Zaffaroni (2000, Remark 4.4)). We set $\sigma^2 = 0.4$.

In analogy to (31) the following parameterization is considered for the nonlinear MA (cf. (9)):

$$1 + \alpha(L;\theta) = (1-L)^{-d}\frac{\chi(L)}{\phi(L)} = (1-L)^{-d}\frac{(1+\chi_1 L + . + \chi_q L^q)}{(1-\phi_1 L - . - \phi_p L^p)}. \qquad (33)$$

Estimation of the nonlinear MA requires $|d| < 1/2$ and the invertibility conditions on $\phi(L), \chi(L)$. In addition to these conditions, estimation of ARCH($\infty$) with parameterization (31) requires also $\psi_i(\theta) \geq 0\,(i \geq 1)$ and $\Sigma(\theta) < \infty$. These conditions impose severe nonlinear restrictions on $\theta$ and are computationally very demanding.

We consider seven time series, three stock indexes returns (S&P 500, FTSE All, FTSE 100) and four foreign exchange rates returns (Yen/US Dollar spot and 1 month forward, US Dollar/Pound spot and 1 month forward). The data are daily, from 1 January 1985 to 1 June 1998, with 3,500 observations (source: Datastream). Returns are calculated as $x_t = \ln(P_t/P_{t-1})$ where $P_t$ denotes the speculative price of the asset. The ACF for the data in levels and in squares are plotted in Figure 1 for lags $1-300$. For all the time series the graphs clearly indicate a marked difference between the ACF in the levels (first and third row) and the ACF in the squares (second and fourth row), the latter exhibiting an approximate hyperbolic behaviour as the lag increases. Table 5.1 reports the results for the LM test $\Phi$, based on $\tau_i = 1/i$ and $\Gamma = \pi^2/6$ (cf. Robinson (1991)). As suggested by the graphs, $\Phi$ is always highly significant for all but the Yen/US Dollar spot series, which is not significant at 1%, providing strong indications of dynamic heteroskedasticity in the data.

As a preliminary exercise, we estimate semiparametrically the parameter of long memory using the local Gaussian estimator of Robinson (1995b). The results, in Table 5.2, clearly support the hypothesis of long memory in squared returns for all the series. Next, the nonlinear MA and ARCH($\infty$) are estimated by the Gaussian estimator. The second and third row of Table 5.2 report the estimates of $d$ for ARCH($\infty$) and nonlinear MA respectively. (Asymptotic standard errors use the estimates of the trispectrum integral proposed by Taniguchi (1982) with a Fejer window.) For the nonlinear MA we also report, in square brackets, the estimates of the parameter $\bar{\kappa}$, the fourth-order cumulant of the standartized residuals. For easy reference, the estimates of the other parameters are not reported. The reported estimates correspond to the parameterization (values of $p$ and $q$) that yields the best performance (with significant parameters estimates), expressed by the degree of white-noisiness of the estimated residuals in levels and squares based on Milhøj (1981) approach.

For all the observed time series of returns there is clear evidence that shocks to conditional volatility do decay with time, but very slowly, in agreement with previous empirical studies (see Ding, Granger, and Engle (1993) and Baillie, Bollerslev, and Mikkelsen (1996) among others). Estimates of $\bar{\kappa}$ show a substancial degree of leptokurtosis, especially for the stock indexes return data. This could explain the smaller values obtained for the semi-parametricestimates of $d$ (see Wright (2000)). Estimating $\bar{\kappa}$ freely rather than equating it a priori to zero should instead attenuate the effect of the negative bias for the nonlinear MA estimates of $d$.

A formal analysis of the goodness-of-fit performance of the two competing models is carried out based on the estimated residuals. The results are reported in Table 5.3, within columns two and five. Milhøj (1981) test statistic is calculated under the hypothesis that the power spectrum of the residuals, in levels and squares, is constant at all frequencies, viz. that the residuals and the residuals squared are white-noise. Under the hypothesis of correct model specification Milhøj (1981) test statistic converges in probability to $1/\pi = 0.318...$ Moreover, the last two columns of Table 5.3 report the values of the test statistic using the raw data, respectively in the levels and the squares, providing a benchmark for the results of the fitted models reported in the first four columns. The results of Table 5.3 indicate a better relative performance of the nonlinear MA in terms of goodness-of-fit with respect to ARCH($\infty$) for all but the stock index S&P 500 and the foreign exchange rate US Dollar/Pound spot data, where the latter model appears superior.

The different performance of the two models is probably caused by the severe parameter restrictions on ARCH($\infty$), in particular $\Sigma(\theta) < \infty$. For instance, when $d = q = 0$, $p = 1$ and $\sigma^2 = 1$ (the ARCH(1)), the latter restriction delivers the well-known condition $0 \leq \phi_1 < 1/\sqrt{3}$. Hence, (29) rules out a lot of possible behaviours for the model ACF. This contrasts with the nonlinear MA case, where this condition is not required and, indeed, one can freely allow $-1 < \phi_1 < 1$ when $d = q = 0$, $p = 1$.

All the computations and the optimizations are based on Gauss and Fortran codes using an IBM RS 6000 and the NAG library optimization routine E04UCF.(All the codes are available on request from the author.) In terms of computational burden, the nonlinear MA involves calculations of the convolutions in the model power spectrum (cf. Theorem 1) whereas for ARCH($\infty$) the burden comes from imposing non-negativity of the $\psi_j(\theta)$ and boundedness of $\Sigma(\theta)$. Therefore, despite the apparently simpler expression for the power spectrum of ARCH($\infty$) (cf. (30)), estimation of the nonlinear MA appears computationally straightforward.

# 7 Conclusions

In this paper we consider the family of nonlinear MA models introduced by Robinson and Zaffaroni (1997) and show that, for a long memory specification, inference of the model can be performed by a frequency domain Gaussian estimator and by a Gaussian pseudo LM test for dynamic conditional heteroskedasticity, establishing the related asymptotic distribution theory. Finally, we have analyzed the effect of allowing for a time-varying conditional mean for the observables when parameterized by stationary ARMA. An interesting development would be a multivariate extension of the nonlinear MA and related inference by a multivariate Gaussian estimator. Further research is needed to accomplish these goals.

# A Second-order properties

**Proof of Theorem 1.**(i) The $x_t - \mu$ are square integrable martingale difference, by Assumption $A$, thus displaying a constant power spectrum. The result follows by simple calculations.
(ii) The ACF of the $y_t$, under Assumption $A$ and $\mu = \rho = 0$ was originally given in RZ, section 4. Let us relax Assumption $A$ setting $\mu_3 = E(\epsilon_t^3)$. For integers $a$, $b$, $c$ such that $a > b \geq c$, set $h_{a,(b,c)} = \rho + \sum_{k=a-b}^{a-c} \alpha_k \epsilon_{t+a-k}$. Then, for any $l > 0$ (without loss of generality),

$$\gamma_{yy}(l) = (i) + (ii) + (iii) + (iv),$$

where

$$\begin{aligned}
(i) &= \mathrm{cov}(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2), \\
(ii) &= \mathrm{cov}(2\mu\epsilon_t h_{0,(-1,-\infty)}, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}), \\
(iii) &= \mathrm{cov}(\epsilon_t^2 h_{0,(-1,-\infty)}^2, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}), \\
(iv) &= \mathrm{cov}(2\mu\epsilon_t h_{0,(-1,-\infty)}, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2).
\end{aligned}$$

Applying the cumulant' theorem (see Leonov and Shiryaev (1959)) to each term and with simple but tedious calculations, for any $l = 0, \pm 1, \ldots$,

$$\begin{aligned}
&\gamma_{yy}(l) = \delta_{\alpha\alpha}(0)(\kappa\sigma^4 + 2\sigma^8)\alpha_{|l|}^2 + 4\rho^2\sigma^6\delta_{\alpha\alpha}(l) + \sigma^4\kappa\delta_{\alpha^2\alpha^2}(l) + 2\sigma^8\delta_{\alpha\alpha}^2(l) \\
&+2\sigma^2\mu\alpha_{|l|}\left[\rho\alpha_{|l|}\mu_3 + 2\rho^2\sigma^2 + 2\sigma^4\delta_{\alpha\alpha}(l)\right] + \rho^2(\kappa\sigma^2 + 2\sigma^6)\alpha_{|l|}^2 \\
&+2\mu_3^2\alpha_{|l|}\delta_{\alpha^2\alpha}(l) + \mu_3\left[2\rho\sigma^4\delta_{\alpha\alpha^2}(l) + 4\rho\sigma^4\alpha_{|l|}\delta_{\alpha\alpha}(l) + 2\rho^3\sigma^2\alpha_{|l|}\right. \\
&\left.+2\rho\sigma^4\alpha_{|l|}\delta_{\alpha\alpha}(0) + 2\rho\sigma^4\delta_{\alpha^2\alpha}(l) + \delta(l,0)\nu_y\right.,
\end{aligned} \tag{35}$$

where

$$\begin{aligned}\nu_y \quad &= (\kappa+2\sigma^4)\left((\rho^2+\sigma^2\delta_{\alpha\alpha}(0))^2+2\sigma^2\delta_{\alpha\alpha}(0)(2\rho^2+\sigma^2\delta_{\alpha\alpha}(0))\right)\\ &+\kappa\left(4\mu_3\rho\delta_{\alpha^2\alpha}(0)+(\kappa+2\sigma^4)\delta_{\alpha^2\alpha^2}(0)\right)\\ &+4\mu_3\mu\left[\rho^3+3\rho\sigma^2\delta_{\alpha\alpha}(0)+\mu_3\delta_{\alpha^2\alpha}(0)\right]+4\rho\sigma^4\mu_3\delta_{\alpha^2\alpha}(0).\end{aligned}$$

Turning to the power spectrum evaluation, under Assumption $A$, using (3) and (4) and setting $k_1 = \sigma^2(\kappa+2\sigma^4)(\rho^2+\sigma^2\delta_{\alpha\alpha}(0))+\sigma^4\kappa\delta_{\alpha\alpha}(0)$, $k_2 = 2\sigma^8$, $k_3 = 4\sigma^6\mu$, $k_4 = 4\sigma^6\rho^2$, $k_5 = 4\sigma^4\rho^2\mu$, $k_6 = \sigma^4\kappa$, yields

$$\begin{aligned}\gamma_{yy}(l) \quad &= \frac{k_1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\beta(\omega)\beta(\lambda)e^{i\omega l}e^{i\lambda l}d\omega d\lambda\\ &+\frac{k_2}{(2\pi)^2}\int_{-\pi}^{\pi}\mid\alpha(\omega)\mid^2\mid\alpha(\lambda)\mid^2 e^{i\omega l}e^{i\lambda l}d\omega d\lambda\\ &+\frac{k_3}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\beta(\lambda)\mid\alpha(\omega)\mid^2 e^{i\omega l}e^{i\lambda l}d\omega d\lambda\\ &+\frac{k_4}{2\pi}\int_{-\pi}^{\pi}\mid\alpha(\lambda)\mid^2 e^{i\lambda l}d\lambda+\frac{k_5}{2\pi}\int_{-\pi}^{\pi}\beta(\lambda)e^{i\lambda l}d\lambda\\ &+\frac{k_6}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\alpha(-\omega-\lambda-\mu)\alpha(\lambda)\alpha(\mu)\alpha(\omega)e^{il(\omega+\lambda)}d\omega d\lambda d\mu.\end{aligned}$$

Then make the change of variable from $\omega$ to $\nu = \omega+\lambda$ and equate the integrand with respect to $\nu$ to $f_{yy}(\nu)e^{il\nu}$. $\square$

# B Asymptotic properties of the Gaussian estimator: martingale difference levels

## B.1 Lemmata

In this section we will establish a number of properties for the nonlinear MA model (needed for the LLN and the CLT results) mainly in terms of the behaviour of the model spectral densities $f_{yy}(\lambda;\psi,\sigma^2)$ and its derivatives. In turn, these are implied by establishing certain asymptotic features of the $\alpha_i(\theta)$ and their convolution. We recall that $\sim$ denotes asymptotic equivalence, $1_{\mathcal{A}}$ denotes the indicator function, $\delta(u,v)$ the Kronecker delta so that $\delta(u,v) = 1_{u=v}$, $tr(.)$ the trace operator and $\parallel . \parallel$ the Euclidean norm. Constants will be denoted $k$ (not always the same) although these will, in general, be a function of the parameter ($k = k(\theta)$) but we will not write this explicitly.

**Lemma 1** *Under Assumption $B_1$, for some integer $m, n$ with $m, n \geq 0$ and $m + n \geq 2$, as $l \to \infty$,*

$$\sum_{j=1}^{\infty} \alpha_j^m(\theta)\, \alpha_{j+l}^n(\theta) \sim k(l^{n(d(\theta)-1)} + l^{(n+m)(d(\theta)-1)+1}),$$

*for some constant $k$.*

**Proof.** As $l \to \infty$, $\sum_{j=1}^{\infty} \alpha_j^m(\theta)\, \alpha_{j+l}^n(\theta) = \sum_{j=1}^{l} \alpha_j^m(\theta)\, \alpha_{j+l}^n(\theta) + \sum_{j=l+1}^{\infty} \alpha_j^m(\theta)\, \alpha_{j+l}^n(\theta)$ $\sim\ k\,(l^{n(d(\theta)-1)} \sum_{j=1}^{l} \alpha_j^m(\theta) + \sum_{j=l+1}^{\infty} \alpha_j^{(m+n)}(\theta)) \sim\ k(l^{n(d(\theta)-1)} + l^{(m+n)(d(\theta)-1)+1})$.$\square$

**Lemma 2** *Let $\{p_j\,, j = 0, \pm 1 \ldots\}$ and $\{q_j\,, j = 0, \pm 1 \ldots\}$ be two square summable sequences satisfying, for $r_j \in \{p_j, q_j\}$,*

$$\begin{aligned} &(a) \quad |\, r_j - r_{j+1}\,| \leq |\, r_j\,|\; k/j, \quad \text{all } j > J, \text{ some } 0 < k, J < \infty,\\ &(b) \quad r_j \to 0, \quad \text{as } j \to \infty. \end{aligned}$$

*Then, the product sequence and the convolution sequence, defined by $m_u = p_u q_u$, $n_u = \sum_{i=-\infty}^{\infty} p_i q_{i+u}$, satisfy (a) and (b). Note that condition (a) is stronger than quasi monotonic convergence to zero (QMC), implying bounded variation as well (Yong 1974, section I.1).*

**Proof.** Let us start from the product sequence $m_j$. From $m_{j+1} - m_j = p_{j+1}(q_{j+1} - q_j) + q_j(p_{j+1} - p_j) = (i) + (ii)$. From the assumptions made one obtains, for some $0 < k, k', J < \infty$ and all $j > J$,
$|(i)| \leq k/j\ |p_{j+1}\, q_j|$, $|(ii)| \leq k'\, 1/j\ |p_j\, q_{j+1}|$, so that as $j \to \infty$,
$|m_j - m_{j+1}| = O(|m_j|\ /j)$. For the convolution, as $j \to \infty$, $n_{j+1} - n_j = \sum_{i=-\infty}^{\infty} p_i(q_{i+j+1} - q_{i+j})$
$= \sum_{|i| \leq j} p_i(q_{i+j+1} - q_{i+j}) + \sum_{|i|>j} p_i(q_{i+j+1} - q_{i+j}) = (iii) + (iv)$.
Then, as $j \to \infty$,
$|(iii)| = O(|\sum_{|i| \leq j}(p_i q_{i+j})/(i+j)\,|) = O(|q_j|\ /j \sum_{|i| \leq j} |\, p_i\,|)$,
$|(iv)| = O(|\sum_{|i|>j} p_i(q_{i+1} - q_i)\,|) = O(1/j \sum_{|i|>j} |\, p_i q_i\,|)$. $\square$

**Lemma 3** *Let the $\delta_{\alpha\alpha}(l;\theta)$ be given by replacing the $\alpha_i$ with the $\alpha_i(\theta)$ in $\delta_{\alpha\alpha}(l)$. By Assumptions $B_1$, $B_2$, the following sequences are QMC:*
$\delta_{\alpha\alpha}(l;\theta)$, $\delta_{\alpha\alpha}^2(l;\theta)$, $\alpha_{|l|}^2(\theta)$, $\alpha_{|l|}(\theta)\delta_{\alpha\alpha}(l;\theta)$, $l = \pm 1, \ldots$
*Assuming further $B_3$, $B_4$ the following are QMC:*
$\frac{\partial}{\partial\theta_i}\delta_{\alpha\alpha}^2(l;\theta)$, $i = 1, 2, .., p$, $l = \pm 1, \ldots$
*Assuming further $B_5$, $B_6$, the following are QMC:*
$\frac{\partial^2}{\partial\theta_{i_1}\partial\theta_{i_2}}\delta_{\alpha\alpha}^2(l;\theta)$, $i_1, i_2 = 1, 2, .., p$, $l = \pm 1, \ldots$

**Proof.** The result follows by the direct use of Lemma 2 in each of the sequences obtained as products or convolutions of the $\alpha_j(\theta)$ and their derivatives. $\square$

**Lemma 4** *Assume $A$, $B_1$, $B_2$. For any $\psi \in \bar{\Psi}$, for constant $0 < k < \infty$,*

$$f_{yy}(\lambda; \psi, \sigma^2) \sim k \mid \lambda \mid^{-2d(\theta)} \quad as\ \lambda \to 0^+.$$

*Let $k$ be a constant $0 <\mid k \mid< \infty$, varying in each sentence below and let $L(\cdot)$ be either constant or equal to $\ln(\cdot)$. Assume further $B_3$, $B_4$. For $j = 1, \ldots, p,$, as $\lambda \to 0^+$ ,*

$$\begin{aligned} \tfrac{\partial}{\partial \theta_j} f_{yy}(\lambda; \psi, \sigma^2) &\sim k\, L(\tfrac{1}{|\lambda|}) \mid \lambda \mid^{-2d(\theta)}, \\ \tfrac{\partial}{\partial \bar{\kappa}} f_{yy}(\lambda; \psi, \sigma^2) &\sim\ k. \end{aligned}$$

*Assume further $B_5$, $B_6$ and let $L_i(\cdot)$ be either constant or equal to $\ln(\cdot)$. For $i_1, i_2 = 1, \ldots, p$, as $\lambda \to 0^+$ ,*

$$\begin{aligned} \tfrac{\partial^2}{\partial \theta_{i_1} \partial \theta_{i_2}} f_{yy}(\lambda; \psi, \sigma^2) &\sim k\, L_{i_1}(\tfrac{1}{|\lambda|})\, L_{i_2}(\tfrac{1}{|\lambda|}) \mid \lambda \mid^{-2d(\theta)}, \\ \tfrac{\partial^2}{\partial \bar{\kappa} \partial \theta_i} f_{yy}(\lambda; \psi, \sigma^2) &\sim\ k. \end{aligned}$$

**Proof.** Let us focus on the model spectrum, which we write as

$$f_{yy}(\nu; \psi, \sigma^2) = f_1(\nu) + f_2(\nu) + f_3(\nu) + f_4(\nu) + \frac{\nu_y(\psi, \sigma^2)}{2\pi},$$

with

$$f_1(\nu) = \tfrac{k_1(\psi,\sigma^2)}{2\pi} \textstyle\sum_{l=-\infty}^{\infty} \alpha^2_{|l|}(\theta) e^{i\nu l}, \qquad f_2(\nu) = \tfrac{k_2(\psi,\sigma^2)}{2\pi} \textstyle\sum_{l=-\infty}^{\infty} \delta^2_{\alpha\alpha}(l; \theta) e^{i\nu l},$$

$$f_3(\nu) = \tfrac{k_3(\psi,\sigma^2)}{2\pi} \textstyle\sum_{l=-\infty}^{\infty} \delta_{\alpha\alpha}(l; \theta) e^{i\nu l}, \quad f_4(\nu) = \tfrac{k_4(\psi,\sigma^2)}{2\pi} \textstyle\sum_{l=-\infty}^{\infty} \delta_{\alpha^2\alpha^2}(l)(\theta) e^{i\nu l},$$

where the $k_1(.,.)$, $k_2(.,.)$, $k_3(.,.)$, $k_4(.,.)$ and $\nu_y(.,.)$ can be easily derived from the proof of Theorem 1 with the $\alpha_i$ replaced by the $\alpha_i(\theta)$. By Lemma 1 and Assumption $B_1$ it follows that the dominating term is $f_4(.)$. Finally, apply Yong (1974, Theorem III-14). The proofs for the first and second derivative of the model spectrum follow along the same lines, where $L(\cdot)$, $L_{i_1}(\cdot)$, $L_{i_2}(\cdot)$ indicate either a constant or $\ln(\cdot)$. □

**Lemma 5** *Under Assumptions $A, B_1, B_2$, $f_{yy}(\lambda; \psi, \sigma^2)$ is continuous for all $\psi$ and $\lambda$ such that $\lambda \neq 0$. Further assuming $B_3$, $B_4$, the same applies to $\frac{\partial}{\partial \psi_i} f_{yy}(\lambda; \psi, \sigma^2)\, (i = 1, 2, .., p+1)$. Further assuming $B_5$, $B_6$, the same applies to $\frac{\partial^2}{\partial \psi_{i_1} \psi_{i_2}} f_{yy}(\lambda; \psi, \sigma^2)\, (i_1, i_2 = 1, 2, .., p+1)$.*

**Proof.** Application of Robinson (1994a, Lemma 8) yields the results. Note that for the derivatives of $f_{yy}(\lambda; \psi, \sigma^2)$ a slowly varying function appears, so that in fact an approximate Lipschitz continuity condition holds. □

**Lemma 6** *Under Assumptions $A$, $B_1$, for any $0 \leq \lambda \leq \pi$,*

$$\inf_{\psi \in \bar{\Psi}} f_{yy}(\lambda; \psi, \sigma^2) \geq \delta > 0,$$

*for some $\delta$ independent from $\psi$.*

**Proof.** For $\lambda \to 0^+$ the result comes directly from Lemma 4. We will then consider the case $0 < \lambda \le \pi$. The result follows decomposing $f_{yy}(\lambda; \psi, \sigma^2)$ as the sum of a non-negative function of $\lambda$ and of a strictly positive quantity, constant in $\lambda$. In fact, from the nonlinear structure of the process $y_t$, the model ACF exhibits a discontinuity (a positive jump) at lag $u = 0$. Given that $\sigma^2 > 0$, we need to show that $h_y(\lambda; \psi)$ is strictly positive. From Theorem 1, for any $l > 0$ (with no loss of generality),

$$\begin{aligned}\bar{\gamma}_{yy}(l;\psi) \quad &= 2(\delta^2_{\alpha\alpha}(l;\theta) - \delta_{\alpha^2\alpha^2}(l;\theta)) + 4\delta_{\alpha\alpha}(l;\theta) + (\bar{\kappa}+2)\delta_{\alpha^2\alpha^2}(l;\theta) \\ &+(2+\bar{\kappa})\alpha_l^2(\theta)(\delta_{\alpha\alpha}(0;\theta)+1), \end{aligned} \tag{36}$$

and, by simple manipulation,

$$\bar{\gamma}_{yy}(0;\psi) = \tilde{\gamma}_{yy}(0;\psi) + \delta(\psi),$$

setting

$$\delta(\psi) = (2+\bar{\kappa})\left(4\delta_{\alpha\alpha}(0;\theta) + 2(\delta^2_{\alpha\alpha}(0;\theta) - \delta_{\alpha^2\alpha^2}(0;\theta))\right) + (2+\bar{\kappa})^2\delta_{\alpha^2\alpha^2}(0;\theta),$$

and

$$\begin{aligned}\tilde{\gamma}_{yy}(0;\psi) \quad &= 2(\delta^2_{\alpha\alpha}(0;\theta) - \delta_{\alpha^2\alpha^2}(0;\theta)) + 4\delta_{\alpha\alpha}(0;\theta) + (\bar{\kappa}+2)\delta_{\alpha^2\alpha^2}(0;\theta) \\ &+(2+\bar{\kappa})(1+\delta_{\alpha\alpha}(0;\theta))^2. \end{aligned} \tag{37}$$

Define another sequence $\tilde{\gamma}_{yy}(l;\psi)$ such that $\tilde{\gamma}_{yy}(l;\psi) = \bar{\gamma}_{yy}(l;\psi)\ (l \neq 0)$ with $\tilde{\gamma}_{yy}(0;\psi)$ defined as above. We need to show only that $\tilde{\gamma}_{yy}(l;\psi)\ (l = 0, \pm 1, \ldots)$ behaves as an autocovariance function by its positive (semi) definiteness and so its Fourier transform is non-negative (see Rozanov (1963, Theorem 5.1)). This property of the $\tilde{\gamma}_{yy}(l;\psi)$ is inherited from the $\bar{\gamma}_{yy}(l;\psi)$ once we show $|\tilde{\gamma}_{yy}(l;\psi)| \le \tilde{\gamma}_{yy}(0;\psi)\ (l = 0, \pm 1, \ldots)$. However, this easily follows comparing (36) and (37) term by term. Hence, the Fourier transform of the sequence $\tilde{\gamma}_{yy}(l;\psi)$ differs from $h_y(\lambda;\psi)$ by $\delta(\psi)$, which is strictly positive for any $\bar{\kappa} > -2$, $d(\theta) > 0$. Finally, set $\delta = \inf_{\psi \in \bar{\Psi}} \delta(\psi)$. $\square$

## B.2 Consistency

In the following three lemmas we will establish the conditions required in order to be able to apply Hannan (1973, Theorem 1). The other two lemmas which follow, used in the proof of Theorem 2, represent the usual steps to establish consistency for M-estimators.

**Lemma 7** *Under Assumption A the process $y_t$ is bounded almost surely, ergodic and strictly stationary.*

**Proof.** We adapt the proof of Nelson (1991, Theorem 2.1). Using Billingsley (1986, Theorem 22.6), the $h_t$, and thus the $y_t$, are finite $a.s.$ given the square summability of the $\alpha_i$ and the independence and finite variance of the $\epsilon_t$, stated in Assumption $A$. Ergodicity and strict stationarity follows by Stout (1974, Theorem 3.5.8). □

**Lemma 8** *Under Assumptions A the process $y_t$ is purely non-deterministic.*

**Proof.** We need to show that (see Hannan (1970))

$$\int_{-\pi}^{\pi} \ln\, f_{yy}(\omega; \psi, \sigma^2) d\omega \quad > -\infty\,.$$

From $|\ln\, f_{yy}(\omega; \psi, \sigma^2)\,| \leq max\left(|\, f_{yy}(\omega; \psi, \sigma^2) - 1\,|, |\,1 - f_{yy}^{-1}(\omega; \psi, \sigma^2)\,|\right)$, the result follows by stationarity of the $y_t$ and continuity of $f_{yy}^{-1}(\lambda; \psi, \sigma^2)$ by Lemma 6. □

**Lemma 9** *Under Assumptions $A, B_1, B_2, B_3, B_4, B_7$, for any $\psi, \psi^* \in \bar{\Psi}$ we obtain $h_y(\lambda; \psi) \neq h_y(\lambda; \psi^*)$, for any $-\pi \leq \lambda < \pi$, whenever $\psi \neq \psi^*$.*

**Proof.** Let us choose two vectors of parameters $\psi, \psi^*$ such that for some $(p+1) \times 1$ vector of constants $\eta \neq 0$ we can write $\psi = \psi^* + \eta$. Let us start, contradicting the conclusion of the Lemma, assuming that for some $-\pi \leq \lambda < \pi$, $h_y(\lambda; \psi) = h_y(\lambda; \psi^*)$. We will see that the condition $\psi \neq \psi^*$ will contradict the hypothesis of the Lemma. By using the fact that the ACF uniquely identifies the power spectrum, it follows that $\bar{\gamma}_{yy}(u; \psi) = \bar{\gamma}_{yy}(u; \psi^*)$, for any $u = \pm 1, \ldots$. In particular, by the mean value theorem, considering the $u_1, ..., u_{p+1}$ of Assumption $B_8$, this implies

$$\begin{pmatrix} \bar{\gamma}_{yy}(u_1; \psi) \\ \vdots \\ \bar{\gamma}_{yy}(u_{p+1}; \psi) \end{pmatrix} = \begin{pmatrix} \bar{\gamma}_{yy}(u_1; \psi) \\ \vdots \\ \bar{\gamma}_{yy}(u_{p+1}; \psi) \end{pmatrix} + \begin{pmatrix} \frac{\partial \bar{\gamma}_{yy}(u_1; \tilde{\psi})}{\partial \psi'} \\ \vdots \\ \frac{\partial \bar{\gamma}_{yy}(u_{p+1}; \tilde{\psi})}{\partial \psi'} \end{pmatrix} (\psi - \psi^*)$$

where $\tilde{\psi}$ is such that $\tilde{\psi} = \psi + R(\psi^* - \psi)$ with $R$ being $(p+1) \times (p+1)$ matrix satisfying $\|\, \tilde{\psi} - \psi \,\| \leq \|\, \psi^* - \psi \,\|$ . However, by Assumption $B_8$, this holds only for $\psi = \psi^*$. □

**Lemma 10**

*(i) Under Assumptions $A, B_1, B_2$, uniformly in $\psi \in \bar{\Psi}$,*

$$\lim_{T \to \infty} Q_T(\psi) = Q(\psi) \;\; a.s.,$$

*with*

$$Q(\psi) = \ln\left(\frac{\sigma^8}{2\pi} \int_{-\pi}^{\pi} \frac{h_y(\lambda)}{h_y(\lambda; \psi)} d\lambda\right) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln\, h_y(\lambda; \psi) + 1.$$

*(ii) Under Assumptions $B_1$, $B_2$, $B_7$, for any $\psi \in \bar{\Psi}$,*

$$Q(\psi) \geq Q(\psi_0) = \ln(\sigma^8) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln\, h_y(\lambda) + 1,$$

*so that the minimum is attained for $\psi = \psi_0$.*

**Proof.** All the convergences below hold as $T \to \infty$. (i) For the first stochastic term on the right hand side of $Q_T(\psi)$ (see 12)), the result follows directly from Hannan (1973, Lemma 1) and continuity of the logarithmic function. Let us consider the second non-stochastic term. Let $j_\epsilon$ be an integer such that $\frac{1}{T} 2\pi j_\epsilon \to \epsilon > 0$ for an arbitrary $0 < \epsilon < 1/2$. Also let $m$ be an integer such that $\frac{1}{m} + \frac{m}{T} \to 0$ so that we can always take $2\pi m/T < \epsilon$ for $T$ large enough. For $h_y(\lambda; \psi)$ being an even function of $\lambda$, one need only consider

$$\frac{1}{T} \sum_{j=1}^{[T/2]} \ln\, h_y(\lambda_j; \psi) = (a) + (b) + (c),$$

with

$$(a) = \frac{1}{T} \sum_{j=j_\epsilon+1}^{[T/2]} \ln\, h_y(\lambda_j; \psi), (b) = \frac{1}{T} \sum_{j=m}^{j_\epsilon} \ln\, h_y(\lambda_j; \psi),\ (c) = \frac{1}{T} \sum_{j=1}^{m-1} \ln\, h_y\,(\lambda_j; \psi).$$

By the continuity of the spectrum away from the zero frequency

$$(a) \to \frac{\int_\epsilon^\pi \ln\, h_y(\lambda; \psi) d\lambda}{2\pi}.$$

Secondly, from $1 - 1/X \leq \ln\, X \leq X - 1$ for $X > 0$, one obtains $(c) = o(1)$.

$$\text{In fact } (c)' = \frac{m}{T} - \frac{1}{T} \sum_{j=1}^{m-1} \frac{1}{h_y(\lambda_j; \psi)} \leq (c) \leq -\frac{m}{T} + \frac{1}{T} \sum_{j=1}^{m-1} h_y\,(\lambda_j; \psi) = (c)''.$$

$(c)' = O(\frac{m}{T}) = o(1)$ by the continuity of the inverse of the spectrum, and $(c)'' = O((\lambda_m)^{2\delta}) = o(1)$ by Robinson (1994b, Lemma 4) where because of $B_1$ we can always choose $\delta > 0$ such that $\sup_{\psi \in \bar{\Psi}} d(\theta) < 1/2 - \delta < 1/2$. For $(b)$ one obtains

$$\mid (b) - \frac{1}{2\pi} \int_0^\epsilon \ln h_y(\omega; \psi) d\omega \mid \leq \mid (b)' \mid + \mid (b)'' \mid,$$

with

$$(b)' = \left( \frac{1}{T} \sum_{j=m}^{j_\epsilon} \ln\, h_y(\lambda_j; \psi) - \frac{1}{2\pi} \int_{\lambda_m}^{\epsilon} \ln\, h_y(\lambda; \psi) d\lambda \right),$$

$$(b)'' = \frac{1}{2\pi} \int_0^{\lambda_m} \ln\, h_y(\lambda; \psi) d\lambda = O(\lambda_m \ln(\lambda_m)).$$

In turn

$$
\begin{aligned}
\mid (b)' \mid \ &\leq \ \mid \frac{\ln\left(h_y(\lambda_m;\psi)\right)}{T} \mid + \frac{1}{2\pi} \mid \sum_{j=m}^{(j_\epsilon)-1} \int_{\lambda_j}^{\lambda_{j+1}} \ln\left(\frac{h_y(\lambda_{j+1};\psi)}{h_y(\lambda;\psi)}\right) d\lambda \mid \\
&= \ \mid (b.1)' \mid + \mid (b.2)' \mid,
\end{aligned}
$$

and $\mid (b.1)' \mid = O(\ln T/T)$. Finally, from

$$
\frac{2\pi}{T} - \frac{\lambda_{j+1}^2 - \lambda_j^2}{2\lambda_{j+1}} \leq \int_{\lambda_j}^{\lambda_{j+1}} \ln\left(\frac{\lambda_{j+1}}{\lambda}\right) d\lambda \leq \lambda_{j+1} \ln\left(\frac{\lambda_{j+1}}{\lambda_j}\right) - \frac{2\pi}{T}
$$

one gets $\frac{\pi}{T(j+1)} \leq \int_{\lambda_j}^{\lambda_{j+1}} \ln(\frac{\lambda_{j+1}}{\lambda}) d\lambda \leq \frac{2\pi}{Tj}$ yielding $\mid (b.2)'' \mid = O(\ln j_\epsilon/T)$, where all the bounds do not depend on $\psi$, by the compactness assumption made on the parameter space.
(ii) $\psi_0$ is the unique minimizer of $Q(\psi)$, given the identification condition, by which $h_y(\lambda;\psi) \neq h_y(\lambda;\psi_0)\,(-\pi < \lambda \leq \pi)$ whenever $\psi \neq \psi_0$ (cf. Lemma 9). Therefore

$$
\begin{aligned}
&Q(\psi) = log(\sigma^8) + log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{h_y(\lambda)}{h_y(\lambda;\psi)} d\lambda\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log h_y(\lambda;\psi) d\lambda + 1 \\
&= log(\sigma^8) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log h_y(\lambda) d\lambda + 1 \\
&+ log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{h_y(\lambda)}{h_y(\lambda;\psi)} d\lambda\right) - \frac{1}{2\pi}\int_{-\pi}^{\pi} log\left(\frac{h_y(\lambda)}{h_y(\lambda;\psi)}\right) d\lambda \\
&= Q(\psi_0) \\
&+ log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{h_y(\lambda)}{h_y(\lambda;\psi)} d\lambda\right) - \frac{1}{2\pi}\int_{-\pi}^{\pi} log\left(\frac{h_y(\lambda)}{h_y(\lambda;\psi)}\right) d\lambda \geq Q(\psi_0),
\end{aligned}
$$

using Jensen's inequality for integrals (see Hardy, Littlewood, and Polya (1964)). □

**Proof of Theorem 2.** By Lemmas 7-10 and Hannan (1973, Lemma 1), strong consistency follows by Hannan (1973, Theorem 1). □

We exploited the ergodicity property for the $a.s.$ convergence of the sample autocovariance. Alternatively, one could use the fourth-order cumulant expression for the $y_t$ in order to derive the convergence in probability of the sample covariances of $y_t$. In fact, one obtains that as $T \to \infty$ $\text{var}(c(u)) = O(T^{2d-1})$, where $c(u)$, $u = 0, \pm 1, \ldots$ denote the sample autocovariances made of $(y_1, \ldots, y_T)'$ (cf. Zaffaroni (1997)).

## B.3 Asymptotic distribution

A number of technical lemmas follow, which allow us to establish the CLT for the score and the convergence of the Hessian. In particular, we establish that the quadratic form $\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\psi_0)I_{yy}(\lambda)d\lambda$, opportunely normalized, is asymptotically normal. This will be derived by studying the asymptotic behaviour of its cumulants of order two and four. The approach here considered is similar to Giraitis and Surgailis (1990, Theorem 1) although there the analysis is based on looking at the power spectrum and trispectrum (the Fourier transform of the fourth-order cumulant) given their simple structure. In our context we need to develop the analysis in terms of the cumulants because, unlike the linear process case of Giraitis and Surgailis (1990), the trispectrum of the nonlinear MA has a very complicated expression (cf. Zaffaroni (1997)).

**Lemma 11** *Under Assumptions $A, B_1, B_2, B_3, B_4, B_5, B_6$, $\int_{-\pi}^{\pi} \ln\, h_y(\omega,\psi)d\omega$ can be differentiated twice under the integral operator for any $\psi \in \bar{\Psi}$.*

**Proof.** We follow Fox and Taqqu (1986) and skip the second differentiation case for simplicity. Denoting the $j$-th unit vector in $R^{p+1}$ by $e_j$, write $\frac{1}{\epsilon}\left(\int_{-\pi}^{\pi} \ln\, h_y(\omega;\psi+e_j\epsilon)d\omega - \int_{-\pi}^{\pi} \ln\, h_y(\omega;\psi)d\omega\right) =$
$\frac{1}{\epsilon}\int_{-\pi}^{\pi} (\ln h_y(\omega;\psi+e_j\epsilon)d\omega - \ln\, h_y(\omega;\psi))\, d\omega$ for some arbitrary $\epsilon > 0$. Thus by the mean value theorem the last integral, ignoring constant terms, is bounded by

$$\int_{-\pi}^{\pi} \mid \omega \mid^{2(d_l-d_u)-\delta} d\omega < \infty,$$

where $d_l = inf_{\theta\in\Theta} d(\theta)$, $d_u = sup_{\theta\in\Theta} d(\theta)$, given $(d_l - d_u) > -1/2$ , where $\delta$ can be taken as arbitrarily small. Use of the dominated convergence theorem concludes. □

**Lemma 12**
*(i) Under Assumptions $A, B_1, B_2, B_7$, the Fourier coefficients of $h_y^{-1}(\lambda;\psi)$ $(\pi \le \lambda < \pi)$ are QMC.*
*(ii) Let $\tilde{g}_\psi(u;\psi)$ $(u = 0, \pm1, ...)$ be the Fourier coefficients of*

$$g(\lambda;\psi) = \frac{\partial}{\partial\psi} h_y^{-1}(\lambda;\psi).$$

*Under Assumptions $A, B_1, B_2, B_3, B_4, B_7$, for any $\psi \in \bar{\Psi}$ and constant $0 <\mid k \mid< \infty$ (varying in each sentence below) the $\tilde{g}_\psi(u;\psi)$ form a vector QMC sequence and, as $u \to \infty$,*

$$\begin{aligned} \tilde{g}_{\theta_i}(u;\psi) &\sim k\, L(|\,u\,|)\; |\,u\,|^{-1-2d(\theta)}, \\ \tilde{g}_{\bar{\kappa}}(u;\psi) &\sim k\;\; |\,u\,|^{-2-6d(\theta)}, \end{aligned}$$

*for $i = 1, .., p$, where the $L(\cdot)$ are either equal to $1$ or $\ln(\cdot)$.*

**Proof.** (i) The continuity and thus the integrability of the inverse of the spectrum yields the convergence to zero of its Fourier coefficients by the Riemann-Lebesgue Theorem. Then the Lipschitz condition (Lemma 5) allows us to use Zygmund (1977, Theorem 4.7-(i) p.46) and so to bound the Fourier coefficients of $h_y^{-1}(\lambda;\psi)$ as $\int_{-\pi}^{\pi} h_y^{-1}(\lambda;\psi)e^{i\lambda u}d\lambda = O(\mid u \mid^{-k})$ for some $0 < k < 1$. By using Yong (1974, Lemma I-1, p.4) the result follows.
(ii) The function $g(\lambda;\psi)$ is given by the product of $h_y^{-2}(\lambda;\psi)$ times $-\frac{\partial}{\partial\psi}h_y(\lambda;\psi)$. Thus, each $\tilde{g}_{\psi_i}(u;\psi)$ is equivalent to the convolution of two series of QMC coefficients by Lemma 3 and part (i) of this Lemma, and therefore it is QMC itself by Lemma 2. By Yong (1974, Theorem III-33-(ii)) and Lemma 4 the asymptotic relations for the Fourier coefficients follow. □

**Lemma 13** *Under Assumptions $A', B$, as $T \to \infty$ ,*

$$T^{\frac{1}{2}}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\psi_0)\left(I_{yy}(\lambda) - EI_{yy}(\lambda)\right)d\lambda\right) \to_d \mathcal{N}_{p+1}(0,\bar{V}),$$

*for some positive semi definite matrix $\bar{V}$.*

**Proof.** All the convergences below hold as $T \to \infty$. We establish the result in two steps. In the first we approximate $\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\psi_0)I_{yy}(\lambda)d\lambda$ by another quadratic form such that they share the same asymptotic distribution. Then, we establish the asymptotic distribution of this second quadratic form, adapting standard limit laws. By Fox and Taqqu (1987, Lemma 8.1), substitution of $\bar{y}$ with $E(y_t)$ in the periodogram is (asymptotically) equivalent. Setting $\tilde{g}_\psi(u) = \tilde{g}_\psi(u;\psi_0)$ $(u = 0, \pm 1, ...)$ and choosing an integer $0 < M < \infty$, define

$$\begin{array}{llll} Q_T & = \sum_{t,s=1}^{T} \tilde{g}_\psi(\mid t-s \mid)\hat{y}_t\hat{y}_s & \text{with} & \hat{y}_t = y_t - E(y_t), \\ S_T & = \sum_{t,s=1}^{T} \tilde{g}_\psi(\mid t-s \mid)\hat{y}_t^{(M)}\hat{y}_s^{(M)} & \text{with} & \hat{y}_t^{(M)} = y_t^{(M)} - E(y_t^{(M)}), \\ \text{and} & & & \\ y_t & = \epsilon_t^2(\rho + \sum_{i=1}^{\infty}\alpha_i\epsilon_{t-i})^2, & E(y_t) & = \sigma^2(\rho^2 + \sigma^2\sum_{i=1}^{\infty}\alpha_i^2), \\ y_t^{(M)} & = \epsilon_t^2(\rho + \sum_{i=1}^{M}\alpha_i\epsilon_{t-i})^2, & E(y_t^{(M)}) & = \sigma^2(\rho^2 + \sigma^2\sum_{i=1}^{M}\alpha_i^2). \end{array}$$

Rather than establishing $\mathrm{var}(Q_T - S_T) = o(T)$ we will obtain the slightly weaker condition: $\mathrm{var}(Q_T - S_T) = O(\eta(M)T)$ where $0 < \eta(M) < \infty$ and $\eta(M) \downarrow 0$ as $M \to \infty$. Then

$$\begin{aligned} \hat{y}_t\hat{y}_s - \hat{y}_t^{(M)}\hat{y}_s^{(M)} &= \left(y_t - E(y_t) - y_t^{(M)} + E(y_t^{(M)})\right)(y_s - E(y_s)) \\ &\quad + \left(y_t^{(M)} - E(y_t^{(M)})\right)\left(y_s - E(y_t) - y_s^{(M)} + E(y_t^{(M)})\right), \\ y_t - y_t^{(M)} &= \epsilon_t^2\Big((\rho + w_{t-1})^2 - (\rho + w_{t-1}^{(M)})^2\Big) \\ &= \epsilon_t^2\left(w_{t-1}^2 - (w_{t-1}^{(M)})^2 + 2\rho(w_{t-1} - w_{t-1}^{(M)})\right) \\ &= \epsilon_t^2\left(2\rho\sum_{i=M}^{\infty}\alpha_i\epsilon_{t-i} + (\sum_{i=M}^{\infty}\alpha_i\epsilon_{t-i})(\sum_{i=1}^{\infty}\bar{\alpha}_i\epsilon_{t-i})\right), \end{aligned}$$

where

$$\bar{\alpha}_i = \begin{cases} \alpha_i, & i \geq M, \\ 2\alpha_i, & i < M. \end{cases}$$

Thus setting $z_t = \left( y_t - y_t^{(M)} - E(y_t) + E(y_t^{(M)}) \right)$ we get $\hat{y}_t \hat{y}_s - \hat{y}_t^{(M)} \hat{y}_s^{(M)} = z_t \hat{y}_s + \hat{y}_t^{(M)} z_s$, yielding

$$\mathrm{var}(Q_T - S_T) = (38.1) + (38.2) + (38.3), \tag{38}$$

with

$$(38.1) = \mathrm{var}(\sum_{t,s=1}^{T} \tilde{g}_\psi(t-s) \hat{y}_t z_s), \quad (38.2) = var(\sum_{t,s=1}^{T} \tilde{g}_\psi(t-s) \hat{y}_t^{(M)} z_s),$$

$$(38.3) = 2 \,\mathrm{cov} \left( \sum_{t_1,s_1=1}^{T} \tilde{g}_\psi(t_1 - s_1) \hat{y}_{t_1} z_{s_1}, \sum_{t_2,s_2=1}^{T} \tilde{g}_\psi(t_2 - s_2) \hat{y}_{t_2}^{(M)} z_{s_2} \right).$$

By the cumulant' theorem (see Brillinger (1975)), denoting by $c_k(a_1, \cdots, a_k)$ the $k$-th order cumulant in possibly $k$ different arguments $a_1, ..., a_k$, the first term of (38) is equivalent to

$$\begin{aligned}
(38.1) \quad &= (a) + (b) + (c), \\
(a) \quad &= \sum_{t_1,s_1,t_2,s_2=1}^{T} \tilde{g}_{\psi_i}(t_1 - s_1) \, \tilde{g}_{\psi_j}(t_2 - s_2) c_2(\hat{y}_{t_1}, \hat{y}_{t_2}) c_2(z_{s_1}, z_{s_2}), \\
(b) \quad &= \sum_{t_1,s_1,t_2,s_2=1}^{T} \tilde{g}_{\psi_i}(t_1 - s_1) \, \tilde{g}_{\psi_j}(t_2 - s_2) c_2(\hat{y}_{t_1}, z_{s_2}) c_2(\hat{y}_{t_2}, z_{s_1}), \\
(c) \quad &= \sum_{t_1,s_1,t_2,s_2=1}^{T} \tilde{g}_{\psi_i}(t_1 - s_1) \, \tilde{g}_{\psi_j}(t_2 - s_2) c_4(\hat{y}_{t_1}, \hat{y}_{t_2}, z_{s_1}, z_{s_2}),
\end{aligned}$$

for $i, j = 1, ..., p+1$. It can be shown that term (38.2) has the same asymptotic behaviour as term (38.1) and both, by Schwarz inequality, determine the behaviour of the covariance term (38.3). Therefore we will develop case (38.1) only. Let us evaluate each of these components. Starting with term $(a)$ above write $z_t = A_t + B_t$ with

$$A_t = 2\rho\epsilon_t^2 (\sum_{i=M}^{\infty} \alpha_i \epsilon_{t-i}), \quad B_t = \epsilon_t^2 (\sum_{i=M}^{\infty} \alpha_i \epsilon_{t-i})(\sum_{i=1}^{\infty} \bar{\alpha}_i \epsilon_{t-i}),$$

and $\hat{y}_t = C_t + D_t + E_t$ with

$$C_t = \rho^2 \epsilon_t^2, \quad D_t = 2\rho\epsilon_t^2 \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i}, \quad E_t = \epsilon_t^2 \sum_{i,j=1}^{\infty} \alpha_j \alpha_i \epsilon_{t-j} \epsilon_{t-i}.$$

Note that each of the $A_t, B_t, D_t, E_t$ is made by products of terms with at least one having mean zero, e.g. $A_t = 2\rho\epsilon_t^2(\sum_{i=M}^{\infty}\alpha_i\epsilon_{t-i})$ and $E(\sum_{i=M}^{\infty}\alpha_i\epsilon_{t-i}) = 0$. Thus we get

$$c_2(z_{s_1}, z_{s_2}) = c_2(A_{s_1}, A_{s_2}) + c_2(B_{s_1}, B_{s_2}) + c_2(A_{s_1}, B_{s_2}) + c_2(B_{s_1}, A_{s_2}), \tag{39}$$

$$\begin{aligned} c_2(\hat{y}_{t_1}, \hat{y}_{t_2}) = &c_2(C_{t_1}, C_{t_2}) + c_2(C_{t_1}, D_{t_2}) + c_2(C_{t_1}, E_{t_2}) \\ &+ c_2(D_{t_1}, C_{t_2}) + c_2(D_{t_1}, D_{t_2}) + c_2(D_{t_1}, E_{t_2}) \\ &+ c_2(E_{t_1}, C_{t_2}) + c_2(E_{t_1}, D_{t_2}) + c_2(E_{t_1}, E_{t_2}). \end{aligned} \tag{40}$$

Obviously there are fewer terms to be evaluated, given that indexes can always be interchanged. Let us set for integers $j, u$

$$\gamma_{j,(u)} = \sum_{i=j}^{\infty}\alpha_i\alpha_{i+u}, \quad \gamma_{(u)} = \gamma_{1,(u)}, \quad \gamma_j = \gamma_{j,(0)}, \quad \gamma = \gamma_{1,(0)} = \gamma_{(0)},$$

$$\delta_{j,(u)} = \sum_{i=j}^{\infty}\alpha_i^2\alpha_{i+u}^2, \quad \delta_{(u)} = \delta_{1,(u)}, \quad \delta_j = \delta_{j,(0)}, \quad \delta = \delta_{1,(0)} = \delta_{(0)}.$$

Note that in terms of the previous definition (cf. section 2) $\gamma_{(u)} = \delta_{\alpha\alpha}(u)$ but we use the newer notation for simplicity. Starting from $c_2(z_{s_1}, z_{s_2})$, using the cumulant' theorem and setting $s = s_2 - s_1$ with $s \geq 0$

$$c_2(A_{s_1}, A_{s_2}) = \begin{cases} (3\sigma^4 + \kappa)\gamma_{M,(s)}, & s > 0, \\ 0, & s = 0, \end{cases}$$

$$c_2(A_{s_1}, B_{s_2}) = c_2(B_{s_1}, A_{s_2}) = 0,$$

$$c_2(B_{s_1}, B_{s_2}) = \begin{cases} \begin{array}{l} \sigma^4\kappa(\alpha_s^2 1_{s\geq M}\gamma_M + \delta_{M,(s)}) + \\ \sigma^8\big(\gamma_{(s)}\gamma_{M,(s)} + \gamma_{M,(s)}(1_{s<M}\gamma_{M-s,(s)} + 1_{s\geq M}\gamma_{M,(s)})\big), \end{array} & s > 0, \\ (\kappa + 2\sigma^4)\sigma^4\gamma_M^2 + (\kappa + 3\sigma^4)(\sigma^4\gamma_M^2 + \kappa\delta_M), & s = 0. \end{cases}$$

$$c_2(C_{s_1}, C_{s_2}) = \begin{cases} \rho^4(\kappa + 2\sigma^4), & s = 0, \\ 0, & s > 0, \end{cases}$$

$$c_2(C_{s_1}, D_{s_2}) = c_2(D_{s_1}, C_{s_2}) = c_2(D_{s_1}, E_{s_2}) = c_2(E_{s_1}, D_{s_2}) = 0,$$

$$c_2(C_{s_1}, E_{s_2}) = \begin{cases} \sigma^2(\kappa + 3\sigma^4), & s = 0, \\ \sigma^2(\kappa + 2\sigma^4)\alpha_s^2, & s > 0, \end{cases}$$

$$c_2(D_{s_1}, D_{s_2}) = \begin{cases} \sigma^2(\kappa + 3\sigma^4)\gamma, & s = 0, \\ \sigma^6\gamma_{(s)}, & s > 0, \end{cases}$$

$$c_2(E_{s_1}, C_{s_2}) = \begin{cases} \sigma^2(\kappa + 3\sigma^4)\gamma, & s = 0, \\ 0, & s > 0, \end{cases}$$

$$c_2(E_{s_1}, E_{s_2}) = \begin{cases} \sigma^4[(\kappa + 2\sigma^4)\gamma\alpha_s^2 + \kappa\delta_{(s)} + 2\sigma^4\gamma_{(s)}^2], & s > 0, \\ \sigma^4[(\kappa + 2\sigma^4)\gamma + \kappa\delta + 2\sigma^4\gamma^2)], & s = 0, \end{cases}$$

For case $(b)$ we no longer have symmetry so we need to take $t > s$, $t = s$, $t < s$ separately, obtaining

$c_2(A_t, C_s) = 0$ for any $t, s$,

$$
\begin{aligned}
&c_2(A_t, D_s) = \begin{cases} 0, & t \neq s, \\ \sigma^2(\kappa + 3\sigma^4)\gamma_M, & t = s, \end{cases} \\
&c_2(A_t, E_s) = 0 \text{ for any } t, s, \\
&c_2(B_t, C_s) = \begin{cases} 0, & t \neq s, \\ \sigma^2(\kappa + 3\sigma^4)\gamma_M, & t = s, \end{cases} \\
&c_2(B_t, D_s) = 0 \text{ for any } t, s, \\
&c_2(B_t, E_s) = \begin{cases} \sigma^4(\gamma(\kappa + 2\sigma^4)\alpha_{t-s}^2 1_{t-s \geq M} + & \\ \kappa\delta_{M-(t-s)+1,(t-s)} + 2\sigma^4\gamma_{M-(t-s)+1,(t-s)}\gamma_{(t-s)}), & t > s, \\ \sigma^2((\kappa + 2\sigma^4)(\gamma + \gamma_M) + \sigma^2(\kappa\delta_M + 2\sigma^4\gamma\gamma_M), & t = s, \\ \sigma^4\left((\kappa + 2\sigma^4)\gamma_M\alpha_{s-t}^2 + \kappa\delta_{M,(s-t)} + 2\sigma^4\gamma_{M,(s-t)}\gamma_{(s-t)}\right), & s > t. \end{cases}
\end{aligned}
$$

As $s \to \infty$, assuming $M < s$, writing $\sum_{i=M}^{\infty} = \sum_{i=M}^{s} + \sum_{i=s+1}^{\infty}$ yields

$$
|\gamma_{M,(s)}| = O(s^{d-1}(s^d + M^d) + s^{2d-1}) = O(s^{2d-1+\delta}M^{-\delta}), \tag{41}
$$

for some arbitrary constant $\delta > 0$. Likewise, when $s < M$

$$
|\gamma_{M,(s)}| = O(\sum_{i=M}^{\infty} \alpha_i^2) = O(M^{-\delta}s^{2d-1+\delta}),
$$

where $0 < \eta(M) = M^{-\delta} < \infty$ can be made arbitrarily small when $M$ gets arbitrarily large. The same bound applies for $\delta_M(s)$. Finally, apply Fox and Taqqu (1987, Theorem 1-(a)), with (using their notation) $p = 2$, $\alpha = 2d$, $\beta = -2d = -\alpha$, yielding $2(\alpha + \beta) = 0 < 1$. Thus $(a) = O(\eta(M)T)$, $(b) = O(\eta(M)T)$.

To evaluate the asymptotic behaviour of the cumulant expression $(c)$ in (38.1) we need to focus on the following possibilities, setting $t = t_1 = t_2$ when $t_1 = t_2$ and $s = s_1 = s_2$ when $s_1 = s_2$ where, by symmetry, we can always take $t_1 \leq t_2$ and $s_1 \leq s_2$:

| | | | | |
|---|---|---|---|---|
| $t = s$, | $t < s$, | $t > s$, | $t < s_1 < s_2$, | $s < t_1 < t_2$, |
| $t = s_1 < s_2$, | $s = t_1 < t_2$, | $s_1 < t < s_2$, | $t_1 < s < t_2$, | $s_1 < t = s_2$, |
| $t_1 < s = t_2$, | $s_1 < s_2 < t$, | $t_1 < t_2 < s$, | $t_1 < t_2 < s_1 < s_2$, | $s_1 < s_2 < t_1 < t_2$, |
| $t_1 < s_1 < t_2 < s_2$, | $s_1 < t_1 < s_2 < t_2$, | $t_1 < t_2 = s_1 < s_2$, | $s_1 < s_2 = t_1 < t_2$, | $t_1 = s_1 < t_2 < s_2$, |
| $s_1 = t_1 < s_2 < t_2$, | $s_1 < t_1 < t_2 < s_2$, | $t_1 < s_1 < s_2 < t_2$, | $s_1 < t_1 < t_2 = s_2$, | $t_1 < s_1 < s_2 = t_2$, |
| $t_1 = s_1 < t_2 = s_2$. | | | | |

We consider the first few cases, given that the same argument will carry through in all cases. We need to evaluate the following fourth-order cumulants:

$$
c_4(x, x', y, y'), \text{ with } x, x' \in \{A_t, B_s\},\ y, y' \in \{C_r, D_u, E_v\},\ \ t, s, r, u, v = 1, \ldots, T
$$

Considering the product expression for $A_t, B_t, C_t, D_t, E_t$, note that when the number of factors involved is odd the cumulant vanishes by $E(\epsilon_t) = 0$. For

example, the cumulant $c_4(A_t, A_s, D_r, E_v)$ involves $2+2+3+4=11$ factors and thus is equal to zero. For the non-zero terms we need to evaluate the cumulant directly.
*Case $s=t$:*
$c_4(A_t, A_t, C_t, C_t) = (c_4(\epsilon_t^2, \epsilon_t^2, \epsilon_t^2, \epsilon_t^2) + 2(c_2(\epsilon_t^2, \epsilon_t^2))^2 + 3\sigma^2 c_3(\epsilon_t^2, \epsilon_t^2, \epsilon_t^2))\, \sigma^2 \gamma_M$, so that $\mid c_4(A_t, A_t, C_t, C_t) \mid \leq K \mid \gamma_M \mid$, where the constant $K$ involves terms such as $\sigma^2$, $\kappa$ and their power. Let us denote by $K_j$, $j=1,2,\ldots$ some constants, (different from time to time) that, like $K$, might involve terms such as $\sigma^2, \kappa, \gamma, \delta$ and their powers. Then

$$
\begin{aligned}
&c_4(A_t, A_t, C_t, E_t) = (K_1\, \gamma_M^2 + K_2 \delta_M),\ c_4(A_t, A_t, D_t, D_t) = (K_1\, \gamma_M^2 + K_2 \delta_M),\\
&c_4(A_t, A_t, E_t, E_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M + K_4 \gamma_M\, \delta_M),\\
&c_4(A_t, B_t, C_t, D_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M),\\
&c_4(A_t, B_t, D_t, E_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M),\\
&c_4(B_t, B_t, C_t, C_t) = (K_1\, \gamma_M^2 + K_2 \delta_M),\\
&c_4(B_t, B_t, C_t, E_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M),\\
&c_4(B_t, B_t, D_t, D_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M),\\
&c_4(B_t, B_t, E_t, E_t) = (K_1\, \gamma_M^2 + K_2 \delta_M + K_3 \gamma_M),
\end{aligned}
$$

and for large $M$

$$
\begin{array}{ll}
\mid c_4(A_t, A_t, C_t, E_t) \mid \leq K \mid \delta_M \mid, & \mid c_4(A_t, A_t, D_t, D_t) \mid \leq K \mid \delta_M \mid,\\
\mid c_4(A_t, A_t, E_t, E_t) \mid \leq K \mid \gamma_M \mid, & \mid c_4(A_t, B_t, C_t, D_t) \mid \leq K \mid \gamma_M \mid,\\
\mid c_4(A_t, B_t, D_t, E_t) \mid \leq K \mid \gamma_M \mid, & \mid c_4(B_t, B_t, C_t, C_t) \mid \leq K \mid \delta_M \mid,\\
\mid c_4(B_t, B_t, C_t, E_t) \mid \leq K \mid \gamma_M \mid, & \mid c_4(B_t, B_t, D_t, D_t) \mid \leq K \mid \gamma_M \mid,\\
\mid c_4(B_t, B_t, E_t, E_t) \mid \leq K \mid \gamma_M \mid . &
\end{array}
$$

*Case $s > t$:*

$$
\begin{aligned}
c_4(A_t, A_t, C_s, E_s) &= \left(K_1\, \alpha_{s-t}^2 \gamma_M + K_2 \alpha_{s-t}^2 \gamma_{M,(s-t)} + K_3 \delta_{M,(s-t)}\right), \\
c_4(A_t, A_t, E_s, E_s) &= \Big(K_1\, \alpha_{s-t}^4 \gamma_M + K_2 \alpha_{s-t}^2 \gamma_{M,(s-t)}^2 + K_3 \alpha_{s-t}^2 \delta_{M,(s-t)} \\
&\quad + K_4 \gamma_{M,(s-t)} \delta_{M,(s-t)} + K_5 \gamma_M + K_6 \gamma_{M,(s-t)}^2 + K_7 \delta_{M,(s-t)}\Big), \\
c_4(A_t, B_t, C_s, D_s) &= \left(K_1\, \delta_{M,(s-t)} + K_2 \gamma_M \gamma_{M,(s-t)} + K_3 \gamma_{M,(s-t)}\right), \\
c_4(A_t, B_t, D_s, E_s) &= \big(K_1\, \alpha_{s-t}^2 \delta_{M,(s-t)} + K_2 \alpha_{s-t}^2 \gamma_{M,(s-t)} \gamma_{(s-t)} + K_3 \alpha_{s-t}^2 \gamma_M \gamma_{(s-t)} \\
&\quad + K_4 \gamma_{M,(s-t)} \delta_{M,(s-t)} + K_5 \gamma_{(s-t)} \delta_{M,(s-t)} + K_6 \delta_{M,(s-t)} \\
&\quad + K_7 \delta_{(s-t)} \gamma_M + K_8 \gamma_M \gamma_{M,(s-t)} + K_9 \gamma_{M,(s-t)}^2 \gamma_{(s-t)}\Big), \\
c_4(B_t, B_t, C_s, E_s) &= \left(K_1\, \delta_M + K_2 \gamma_M^2 + K_3 \delta_{M,(s-t)} + K_4 \gamma_{M,(s-t)}^2\right), \\
c_4(B_t, B_t, D_s, D_s) &= \left(K_1\, \delta_M + K_2 \gamma_M^2 + K_3 \delta_{M,(s-t)} + K_4 \gamma_{M,(s-t)}^2\right), \\
c_4(B_t, B_t, E_s, E_s) &= \big(\alpha_{s-t}^4 (K_1 \delta_M + K_2 \gamma_M^2 + K_3 \gamma_M) + \alpha_{s-t}^2 (K_4 \delta_{(s-t)} \gamma_M \\
&\quad + K_5 \delta_{M,(s-t)} \gamma_M + K_6 \delta_{M,(s-t)} \gamma_{M,(s-t)} + K_7 \delta_{M,(s-t)} \gamma_{(s-t)}) \\
&\quad + K_8 \gamma_{M,(s-t)} \gamma_M + K_9 \gamma_M + K_{10} \delta_M + K_{11} \gamma_M^2 + K_{12} \gamma_{M,(s-t)}^2 \\
&\quad + K_{13} \delta_{M,(s-t)} + K_{14} \gamma_M \gamma_{(s-t)} \gamma_{M,(s-t)} \\
&\quad + K_{15} \gamma_M \delta_{M,(s-t)} + K_{16} \gamma_M \delta_{M,(s-t)} \gamma_{M,(s-t)} \\
&\quad + K_{17} \gamma_M \delta_{(s-t)} + K_{18} \gamma_M \gamma_{(s-t)}^2\big).
\end{aligned}
$$

So, given that as $u \to \infty$, $\alpha_u^m = o(\gamma_{(u)})$, $m \geq 1$ and $\delta_{(u)} = o(\gamma_{(u)})$, one obtains as $(s-t) \to \infty$ and for large $M$

$$
\begin{aligned}
&\mid c_4(A_t, A_t, C_s, E_s) \mid \leq K \alpha_{s-t}^2 \mid \gamma_{M,(s-t)} \mid, && \mid c_4(A_t, A_t, E_s, E_s) \mid \leq K \mid \gamma_M \mid, \\
&\mid c_4(A_t, B_t, C_s, D_s) \mid \leq K \mid \gamma_{M,(s-t)} \mid, && \mid c_4(A_t, B_t, D_s, E_s) \mid \leq K \mid \delta_{M,(s-t)} \mid, \\
&\mid c_4(B_t, B_t, C_s, E_s) \mid \leq K \mid \delta_M \mid, && \mid c_4(B_t, B_t, D_s, D_s) \mid \leq K \mid \delta_M \mid, \\
&\mid c_4(B_t, B_t, E_s, E_s) \mid \leq K \mid \gamma_M \mid.
\end{aligned}
$$

The other cases follow along the same lines, so that by (41) and applying Fox and Taqqu (1987, Theorem 1 a)) $(c) = O(\eta(M)\, T)$, concluding the main part of the proof. Finally, given that $S_T$ is a quadratic form in $M$-dependent variates and thus $\phi$-mixing with arbitrarily fast decreasing mixing coefficient, Ibragimov and Linnik (1971, Theorem 18.5.2) applies. □

**Lemma 14** *Under Assumptions $A, B_1, B_2, B_3, B_4, B_5, B_6$, as $T \to \infty$,*

$$
M_T(\psi) = \frac{\partial^2}{\partial \psi \partial \psi'} Q_T(\psi) \to_{a.s.} M(\psi),
$$

*uniformly in $\psi \in \bar{\Psi}$.*

**Proof.** All the convergences below hold as $T \to \infty$. $M_T(\psi) = (a) + (b) + (c) + (d)$, with

$$(a) = 1/\left(\frac{1}{T}\sum_j \frac{I_{yy}(\lambda_j)}{h_y(\lambda_j;\psi)}\right)\frac{1}{T}\sum_j \frac{\partial^2}{\partial\psi\partial\psi'}h_y^{-1}(\lambda_j;\psi)I_{yy}(\lambda_j),$$

$$(b) = -1/\left(\frac{1}{T}\sum_j \frac{I_{yy}(\lambda_j)}{h_y(\lambda_j;\psi)}\right)^2\left(\frac{1}{T}\sum_j \frac{\partial}{\partial\psi}h_y^{-1}(\lambda_j;\psi)I_{yy}(\lambda_j)\right)$$

$$\times\left(\frac{1}{T}\sum_j \frac{\partial}{\partial\psi'}h_y^{-1}(\lambda_j;\psi)I_{yy}(\lambda_j)\right),$$

$$(c) = \frac{1}{T}\sum_j \frac{\partial}{\partial\psi}h_y^{-1}(\lambda_j;\psi)\frac{\partial}{\partial\psi'}h_y(\lambda_j;\psi),$$

$$(d) = \frac{1}{T}\sum_j h_y^{-1}(\lambda_j;\psi)\frac{\partial^2}{\partial\psi\partial\psi'}h_y(\lambda_j;\psi),$$

where the index $j$ in the summations ranges over $1 \leq j \leq T-1$. Now, for $(a)$ and $(b)$ by Lemma 5 and following the argument of Hannan (1973, Lemma 1)

$(a) \rightarrow_{a.s.} \frac{1}{\left(\int_{-\pi}^{\pi}\frac{h_y(\lambda)}{h_y(\lambda;\psi)}d\lambda\right)}\left(\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\psi\partial\psi'}h_y^{-1}(\lambda;\psi)\,h_y(\lambda)d\lambda\right),$

$(b) \rightarrow_{a.s} -\frac{1}{\left(\int_{-\pi}^{\pi}\frac{h_y(\lambda)}{h_y(\lambda;\psi)}\right)^2}\left(\int_{-\pi}^{\pi}g(\lambda;\psi)h_y(\lambda)d\lambda\right)\left(\int_{-\pi}^{\pi}g'(\lambda;\psi)h_y(\lambda)d\lambda\right).$

The last two terms, $(c)$ and $(d)$, though not stochastic, need further attention owing to the behaviour near zero frequency of the model power spectrum. We consider the proof for $(c)$ only, the result being valid for $(d)$ as well given that they display the same behaviour in the neighborhood of the zero frequency (cf. Lemma 4). Following the proof of Lemma 10, let us consider

$$\frac{1}{T}\sum_{j=1}^{|T/2|}\frac{\partial h_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial h_y(\lambda_j;\psi)}{\partial\psi'} = \frac{1}{T}\sum_{j=1}^{m-1}\frac{\partial h_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial h_y(\lambda_j;\psi)}{\partial\psi'}$$

$$+\frac{1}{T}\sum_{j=m}^{j_\epsilon}\frac{\partial h_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial h_y(\lambda_j;\psi)}{\partial\psi'} + \frac{1}{T}\sum_{j=j_\epsilon+1}^{|T/2|}\frac{\partial h_y(\lambda_j;\psi)^{-1}}{\partial\psi}\frac{\partial h_y(\lambda_j;\psi)}{\partial\psi'}$$

$$= (c.1) + (c.2) + (c.3).$$

By Lemma 5

$$(c.3) \rightarrow \frac{1}{2\pi}\int_\epsilon^\pi \frac{\partial h_y(\lambda;\psi)^{-1}}{\partial\psi}\frac{\partial h_y(\lambda;\psi)}{\partial\psi'}d\lambda.$$

For any $i,j = 1,2,\ldots,(p+1)$, by Lemma 4, as $\lambda \rightarrow 0^+$,

$$\left|\frac{\partial h_y^{-1}(\lambda;\psi)}{\partial\psi_i}\frac{\partial h_y(\lambda;\psi)}{\partial\psi_j}\right| = O(\ln^2(\frac{1}{\lambda})),$$

yielding $|(c.1)| = O(\sum_{j=1}^{m-1} \ln^2(\lambda_j)/T) = O(\ln^2(\lambda_m)\lambda_m) = o(1)$. Finally consider

$$|(c.2)| \leq |(c.2)'| \ + \ |(c.2)''|,$$

setting

$$(c.2)' = \frac{\sum_{j=m}^{j_\epsilon} \ln^2(\lambda_j)}{T} - \frac{\int_{\lambda_m}^{\epsilon} \ln^2(\lambda) d\lambda}{2\pi}, \ (c.2)'' = \int_0^{\lambda_m} \ln^2(\lambda) d\lambda.$$

By solving the integral $|(c.2)''| = O(\lambda_m \ln^2(\lambda_m)) = o(1)$ and $|(c.2)'| = o(1)$ given

$$|(c.2)'| \leq |\frac{\ln^2(\lambda_m)}{T}| + \frac{1}{2\pi} \sum_{j=m}^{j_\epsilon - 1} \int_{\lambda_j}^{\lambda_{j+1}} |\left(\ln^2(\lambda_{j+1}) \ - \ \ln^2(\lambda)\right)| \, d\lambda,$$

and bounding the second term on the right hand side above as

$$|\sum_{j=m}^{j_\epsilon-1} \int_{\lambda_j}^{\lambda_{j+1}} \left(\ln^2(\lambda_{j+1}) \ - \ \ln^2(\lambda)\right) d\lambda \ | \leq | \ \frac{2\pi}{T} \sum_{j=m}^{j_\epsilon-1} \left(\ln^2(\lambda_{j+1}) \ - \ \ln^2(\lambda_j)\right) |$$
$$= \frac{2\pi}{T} \ | \ \left(\ln^2(\lambda_{j_\epsilon-1}) - \ln^2(\lambda_m)\right) | = \frac{2\pi}{T} \ | \ (\ln(j_\epsilon/m) \ln(\lambda_{j_\epsilon+m-1})) \ | = O(\frac{\ln(j_\epsilon)}{T}),$$

where all the bounds are independent of $\psi$. □

**Lemma 15** *Assume* $A', B_1, B_2, B_3, B_4$.
*(i) Setting* $g(\lambda) = g(\lambda; \psi_0)$,

$$\lim_{T\to\infty} T \ \text{var}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda) I_{yy}(\lambda) d\lambda\right)$$
$$= \frac{1}{\pi}\int_{-\pi}^{\pi} g(\omega) g'(\omega) f_{yy}^2(\omega) d\omega$$
$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1) g'(\omega_2) Q_{yyy}(-\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2.$$

*(ii) As* $T \to \infty$

$$E\left(\frac{1}{\sigma^8 2\pi}\int_{-\pi}^{\pi} g(\lambda) I_{yy}(\lambda) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\psi} \ln \ h_y(\lambda) d\lambda\right) = o(T^{-\frac{1}{2}}). \qquad (42)$$

**Proof.** (i) Given the non-Gaussianity of the process, one obtains the two covariance terms and the fourth-order cumulant term:

$$\text{var}(\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda) I_{yy}(\lambda) d\lambda) = (a) + (b),$$

$$(a) \qquad = \frac{1}{T^2(2\pi)^2} \sum_{\substack{t_1,s_1,t_2,s_2\\=1}}^{T} \tilde{g}_{\psi}(t_1-s_1)\tilde{g}'_{\psi}(t_2-s_2)$$
$$\times\left(\gamma_{yy}(t_1-s_2)\gamma_{yy}(t_2-s_1)+\gamma_{yy}(t_1-t_2)\gamma_{yy}(s_2-s_1)\right)$$
$$(b) \qquad = \frac{1}{T^2(2\pi)^2} \sum_{\substack{t_1,s_1,t_2,s_2\\=1}}^{T} \tilde{g}_{\psi}(t_1-s_1)\tilde{g}'_{\psi}(t_2-s_2) cum_{yyy}(s_1-t_1,t_2-t_1,s_2-t_1),$$

where $cum_{yyy}(.,.,.)$ denotes the fourth order cumulant of $y_t$. With respect to (a) the result follows by Fox and Taqqu (1987, Theorem 1-(i)), setting $\alpha=-\beta$ and with respect to (b) by adapting Hosoya (1993, Lemma A4.2) with simple yet tedious calculations.
(ii) From $\int_{-\pi}^{\pi} w(\lambda;\psi)d\lambda = -\int_{-\pi}^{\pi} g(\lambda;\psi)h_y(\lambda;\psi)d\lambda$, by Parseval relation and re-arranging terms, the left hand side of (42) is equal to

$$-2\left(\sum_{u=T-1}^{\infty} \tilde{g}_{\psi}(u)\,\gamma_{yy}(u)\; + \sum_{u=1}^{T-1} \frac{u}{T}\tilde{g}_{\psi}(u)\gamma_{yy}(u)\right) \sim k\,L(T)\,T^{-1} \text{ as } T\to\infty,$$

using (7) and Lemma 12-(ii) where $|L(T)|=O(\ln(T))$. $\square$
**Proof of Theorem 3.** Assuming that the true value $\psi_0$ lies in the interior of $\Psi$ and given the consistency of $\hat{\psi}_T$, by Lemma 11 and using the Delta method

$$\frac{\partial Q_T(\hat{\psi}_T)}{\partial\psi} = 0 = \frac{\partial Q_T(\psi_0)}{\partial\psi} + \frac{\partial^2 Q_T(\tilde{\psi})}{\partial\psi\partial\psi'}\,(\hat{\psi}_T-\psi_0),$$

where $\tilde{\psi}$ is such that $\tilde{\psi}=\psi_0+R(\hat{\psi}_T-\psi_0)$ with $R$ being $(p+1)\times(p+1)$ matrix such that $\parallel \tilde{\psi}-\psi_0 \parallel \leq \parallel \hat{\psi}_T-\psi_0 \parallel$. The result follows combining Lemmas 11-15, yielding for $T\to\infty$

$$M_T(\psi) \to_{a.s.} M(\psi)\,, \text{ uniformly for } \psi\in\bar{\Psi},$$
$$T^{\frac{1}{2}}\frac{\partial Q_T(\psi_0)}{\partial\psi} \to_d \mathcal{N}_{p+1}(0,V). \quad \square$$

# C Asymptotic properties of the Gaussian estimator: ARMA levels

In this section we show how to adapt the results in appendixes B.2 and B.3 for estimation of the nonlinear MA parameters $\psi$, using the squared estimated residuals obtained from preliminary estimation of the ARMA parameters $\zeta$.

At first, we need to establish the asymptotic properties of the Whittle estimator of the ARMA parameters.
**Proof of Theorem 5.** This follows adapting Hannan (1973) (see also Brockwell and Davis (1987, section 10.8) for more details). For $1+\delta_{\alpha\alpha}(0)>0$ and

under Assumption $A$, the $u_t$ are white noise not degenerate, and the strong consistency proof of Hannan (1973, Theorem 1) applies. For the asymptotic distribution one gets:

$$\left[-\frac{\partial^2 P_T(\tilde{\zeta})}{\partial\zeta\partial\zeta'}\right] T^{\frac{1}{2}}(\hat{\zeta}_T - \zeta_0) = T^{\frac{1}{2}}\frac{\partial P_T(\zeta_0)}{\partial\zeta},$$

for $T$ sufficiently large such that the matrix in square brackets is nonsingular, where $\tilde{\zeta}$ satisfies $\parallel \tilde{\zeta} - \zeta_0 \parallel \leq \parallel \hat{\zeta}_T - \zeta_0 \parallel$ .

All the convergences below hold as $T \to \infty$. By the same arguments of Hannan (1973, Lemma 1) and simple manipulations

$$\frac{\partial^2 P_T(\tilde{\zeta})}{\partial\zeta\partial\zeta'} \to_{a.s} \gamma_{uu}(0)\, F,$$

with $F$ defined in (22). Next, we need to show that

$$T^{\frac{1}{2}}\frac{\partial P_T(\zeta_0)}{\partial\zeta} \to_d N_{r+q}(0, \gamma_{uu}^2(0)\, G),$$

with $G$ defined in (23). Exploiting the martingale difference property of the $u_t$ (see Hannan (1973, Lemma 5)) we just need to establish the CLT for

$$\frac{2\pi}{T^{\frac{1}{2}}}\sum_{j=1}^{T-1} I_{uu}(\lambda_j) b(\lambda_j),$$

where $b(\lambda)$ is defined in (24). The $u_t$ are not conditionally homoskedastic, violating (6) of Hannan (1973)). However, by a truncation argument precisely along the lines of Lemma 13, we approximate the $u_t$ with $u_t^{(N)} = \epsilon_t(\rho + \sum_{j=1}^{N} \alpha_j \epsilon_{t-j})$ for some integer $N < \infty$. By the $N$-dependence of the $u_t^{(N)}$

$$\frac{2\pi}{T^{\frac{1}{2}}}\sum_{j=1}^{T-1} I_{u^{(N)}u^{(N)}}(\lambda_j) b(\lambda_j) \to_d N_{r+q}(0, \bar{V}),$$

for some positive semi definite matrix $\bar{V} = \bar{V}(N)$ and, with simple yet tedious calculations,

$$\operatorname{var}\left(\frac{2\pi}{T^{\frac{1}{2}}}\sum_{j=1}^{T-1} (I_{u^{(N)}u^{(N)}}(\lambda_j) - I_{uu}(\lambda_j)) b(\lambda_j)\right) = O(\eta(N)),$$

for some function $0 < \eta(N) < \infty$ with $\eta(N) \downarrow 0$ as $N \to \infty$. Note, though, that the special structure (2) of the $u_t$ implies a form of the asymptotic variance-covariance matrix of the $\hat{\zeta}_T$ that is different from the conditionally

homoskedastic case of Hannan (1973). In fact, setting
$\tilde{b}_\zeta(u) = 1/2\pi \int_{-\pi}^{\pi} b(\lambda)e^{-iu\lambda}d\lambda \, (u = 0, \pm 1, ...)$,

$$\begin{aligned}
&\operatorname{var}\left(\frac{2\pi}{T^{\frac{1}{2}}}\sum_{j=1}^{T-1} I_{uu}(\lambda_j)b(\lambda_j)\right) \\
&= \frac{1}{T}\sum_{h_1,h_2=-\infty}^{\infty} \tilde{b}_\zeta(h_1)\tilde{b}_\zeta(h_2)' \sum_{s_1=1}^{T-|h_1|}\sum_{s_2=1}^{T-|h_2|} E(u_{s_1}u_{s_1+h_1}u_{s_2}u_{s_2+h_2}) + o(1) \\
&= 2\sum_{h=-T+1}^{T-1} \tilde{b}_\zeta(h)\tilde{b}_\zeta(h)'(1 - \frac{|\, h \,|}{T})E(u_s^2 u_{s+h}^2) + o(1) \\
&= 2\gamma_{uu}^2(0)\sum_{h=-T+1}^{T-1}(1 - \frac{|\, h \,|}{T})\tilde{b}_\zeta(h)\tilde{b}_\zeta(h)' \\
&+2\sum_{h=-T+1}^{T-1}(1 - \frac{|\, h \,|}{T})\tilde{b}_\zeta(h)\tilde{b}_\zeta(h)'\gamma_{u^2u^2}(h) + o(1),
\end{aligned}$$

given that the contribution of $E(u_{s_1}u_{s_1+h_1}u_{s_2}u_{s_2+h_2})$ is non zero only for cases $s_1 = s_2$, $h_1 = h_2$ and $s_1 = s_2 + h_2$, $s_2 = s_1 + h_1$ using $1/2\pi \int_{-\pi}^{\pi} b(\lambda)d\lambda = \tilde{b}_\zeta(0) = 0$ (vector of zeros) and $\tilde{b}_\zeta(-h) = \tilde{b}_\zeta(h)$. The last equality above is obtained writing $E(u_s^2 u_{s+h}^2) = \gamma_{uu}^2(0) + \gamma_{u^2u^2}(h)$. The standard case of independent $u_t$, or more generally with conditional homoskedasticity is obtained when $\gamma_{u^2u^2}(u) = 0$ for $u \neq 0$.

Thus,

$$\begin{aligned}
&\operatorname{var}\left(\frac{2\pi}{T^{\frac{1}{2}}}\sum_{j=1}^{T-1} I_{uu}(\lambda_j)b(\lambda_j)\right) \to 2\gamma_{uu}^2(0)\sum_{h=-\infty}^{\infty}\tilde{b}_\zeta(h)\tilde{b}_\zeta(h)' \\
&+2\sum_{h=-\infty}^{\infty}\tilde{b}_\zeta(h)\tilde{b}_\zeta(h)'\gamma_{u^2u^2}(h) \\
&= \frac{\gamma_{uu}^2(0)}{\pi}\int_{-\pi}^{\pi} b(\lambda)b(\lambda)'d\lambda + \frac{1}{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{u^2u^2}(\lambda_1)b(\lambda_2)b(\lambda_1-\lambda_2)'d\lambda_1 d\lambda_2. \quad \square
\end{aligned}$$

**Proof of Theorem 6.** All the limits below are taken for $T \to \infty$. For consistency, we just need to show that

$$\sup_{\psi \in \bar{\Psi}} |\, Q_T(\psi) - \tilde{Q}_T(\psi) \,| \to_{a.s} 0, \tag{43}$$

so that Theorem 2 applies. For this purpose, given the smoothness of the $\pi_j(\zeta)$ and using the mean value theorem

$$\hat{u}_t \quad = u_t + \sum_{j=1}^{t-1}(\pi_j(\zeta_0) - \pi_j(\hat{\zeta}_T))x_{t-j} + \sum_{j=t}^{\infty}\pi_j(\zeta_0)x_{t-j}$$

$$= u_t + (\zeta_0 - \hat{\zeta}_T)'\tilde{v}_t + B_t,$$

setting $\tilde{v}_t = \sum_{j=1}^{t-1} \vartheta_j(\tilde{\zeta}) x_{t-j}$, for some $\tilde{\zeta}$ satisfying $\| \tilde{\zeta} - \zeta_0 \| \leq \| \hat{\zeta}_T - \zeta_0 \|$, and $B_t = \sum_{j=0}^{\infty} \pi_{j+t}(\zeta_0) x_{-j}$. Thus

$$\hat{u}_t^2 = u_t^2 + ((\zeta_0 - \hat{\zeta}_T)'\tilde{v}_t)^2 + B_t^2 + 2u_t(\zeta_0 - \hat{\zeta}_T)'\tilde{v}_t + 2B_t(\zeta_0 - \hat{\zeta}_T)'\tilde{v}_t + 2u_t B_t. \quad (44)$$

By Theorem 5, (44) and $|B_t| = O(a^t)$, $t \to \infty$ $a.s.$ for some $0 < a < 1$ one gets

$$I_{\hat{u}^2\hat{u}^2}(\lambda) = I_{u^2u^2}(\lambda) + (\zeta_0 - \hat{\zeta}_T)' 4\mathcal{R}e(I_{u^2\,u\tilde{v}}(\lambda)) + o(T^{-\frac{1}{2}}) R_T(\lambda) \quad a.s., \ -\pi \leq \lambda < \pi, \quad (45)$$

for some real random variable $R_T(\lambda)\,(-\pi \leq \lambda < \pi)$, function of the $u_t$, $B_t$ and $\tilde{v}_t\,(1 \leq t \leq T)$ satisfying, by the same arguments of Hannan (1973, Lemma 1),

$$\frac{2\pi}{T} \sum_{j=1}^{T-1} R_T(\lambda) \to_{a.s.} \int_{-\pi}^{\pi} R(\lambda) d\lambda,$$

for some real integrable (by Assumption $A'$) function $R(\lambda)$. Likewise,

$$\frac{4}{T} \sum_{j=1}^{T-1} \mathcal{R}e(I_{u^2\,u\tilde{v}}(\lambda_j)) \to_{a.s.} \frac{2}{\pi} \int_{-\pi}^{\pi} \mathcal{R}e(f_{u^2\,uv}(\lambda)) d\lambda,$$

with a finite limit (by Assumption $A'$), where we recall that, given $\lambda_j = 2\pi/T$,

$$I_{u^2\,u\tilde{v}}(\lambda_j) = \frac{1}{2\pi T} \sum_{t,s=1}^{T-1} u_t^2 u_s \tilde{v}_s e^{i(t-s)\lambda_j}, \ \ 1 \leq j \leq T-1.$$

Using $\ln(1 + a/b) \leq a/b$ for any positive $a, b$, and (45), Lemma 6 and the consistency part of Theorem 5 then

$$\begin{aligned}
&| Q_T(\psi) - \tilde{Q}_T(\psi) | \\
&\leq \frac{|\zeta_0 - \hat{\zeta}_T|'}{\hat{\sigma}_T^8(\psi)} \frac{4}{T} \sum_{j=1}^{T-1} \frac{|\mathcal{R}e(I_{u^2\,u\tilde{v}}(\lambda_j))|}{h_y(\lambda_j;\psi)} + \frac{o(|\zeta_0 - \hat{\zeta}_T|')}{\hat{\sigma}_T^8(\psi)} \frac{\frac{1}{T}\sum_{j=1}^{T-1} |R_T(\lambda_j)|}{h_y(\lambda_j;\psi)} \\
&\leq \frac{1}{\delta\,\hat{\sigma}_T^8(\psi)} o\left( \frac{4}{T} \sum_{j=1}^{T-1} |\mathcal{R}e(I_{u^2\,u\tilde{v}}(\lambda_j))| + \frac{1}{T} \sum_{j=1}^{T-1} |R_T(\lambda_j)| \right) \quad a.s.,
\end{aligned}$$

where the term in square brackets does not depend on $\psi$ and $\inf_{\psi \in \bar{\Psi}} \hat{\sigma}_T^8(\psi) > 0$, given $\sup_{\psi \in \bar{\Psi}} h_y(\lambda; \Psi) < \infty$ for $0 < \lambda \leq \pi$ by Lemma 5.

For the asymptotic distribution result we just need to look at the asymptotic distribution of the normalized score evaluated at $\psi_0$, which satisfies

$$\begin{aligned}
&T^{\frac{1}{2}}\frac{\partial \tilde{Q}_T(\psi_0)}{\partial\psi} \\
&= \frac{\hat{\sigma}^8_T(\psi_0)}{\tilde{\sigma}^8_T(\psi_0)}T^{\frac{1}{2}}\frac{\partial Q_T(\psi_0)}{\partial\psi} \\
&+\frac{1}{\tilde{\sigma}^8_T(\psi_0)}\frac{4}{T}\sum_{j=1}^{T-1} g(\lambda_j;\psi_0)\mathcal{R}e(I_{u^2\,u\tilde{v}'}(\lambda_j))\,T^{\frac{1}{2}}(\zeta_0-\hat{\zeta}_T)+o(1)\quad a.s.
\end{aligned}$$

by (45). Given that

$$\frac{4}{T}\sum_{j=1}^{T-1} g(\lambda_j;\psi_0)\mathcal{R}e(I_{u^2\,u\tilde{v}'}(\lambda_j)) \to_{a.s.} H(\psi_0),\ \ \frac{\hat{\sigma}^8_T(\psi_0)}{\tilde{\sigma}^8_T(\psi_0)} \to_{a.s.} 1,$$

the normalized score is asymptotically equivalent to

$$T^{\frac{1}{2}}\frac{\partial Q_T(\psi_0)}{\partial\psi} +\ H(\psi_0)\,T^{\frac{1}{2}}(\zeta_0-\hat{\zeta}_T)+o(1),\ \ a.s.$$

By Theorem 3 and 5 then

$$T^{\frac{1}{2}}\frac{\partial Q_T(\psi_0)}{\partial\psi} \to_d \mathcal{N}_{p+1}(0, V+H(\psi_0)F^{-1}GF^{-1}H(\psi_0)'+H(\psi_0)Z+Z'H(\psi_0)'),$$

setting

$$Z=\lim_{T\to\infty} T\mathrm{cov}\left((\zeta_0-\hat{\zeta}_T), \frac{\partial Q_T(\psi_0)}{\partial\psi'}\right).$$

It turns out that $Z=0$ under Assumption $A'$. In fact, we just need to look at

$$\begin{aligned}
&T\mathrm{cov}\left(\frac{1}{T}\sum_{j_1=1}^{T-1} I_{uu}(\lambda_{j_1})b(\lambda_{j_1}), \frac{1}{T}\sum_{j_2=1}^{T-1} I_{u^2u^2}(\lambda_{j_2})g(\lambda_{j_2})'\right) \\
&= \frac{1}{T}\sum_{t_1,t_2,s_1,s_2=1}^{T} \tilde{b}_\zeta(t_1-t_2)\tilde{g}'_\psi(s_1-s_2)\mathrm{cov}(u_{t_1}u_{t_2}, u^2_{s_1}u^2_{s_2})+o(1).
\end{aligned}$$

Given that $\tilde{b}_\zeta(0)=0$ the only non zero terms would be $\tilde{b}_\zeta(t-s)\,\tilde{g}'_\psi(t-s)E(u^3_tu^3_s)\,(1\le t\ne s\le T)$ and $\tilde{b}_\zeta(t-s)\,\tilde{g}'_\psi(0)E(u^5_tu_s)\,(1\le s<t\le T)$. However, $E(\epsilon^3_t)=E(\epsilon^5_t)=0$ by Assumption $A'$ yielding $E(u^3_tu^3_s)=E(u^5_tu_s)=0$ for $t\ne s$. $\square$

# D Asymptotic distribution of the LM test under $H_0$

In this section we derive the asymptotic distribution of the LM test statistics $\Phi$ (cf. (14)) under $H_0$. Finally, we will introduce a robustified LM test statistic and show that it shares the same asymptotic distribution of $\Phi$ under $H_0$. By Assumption $A$

$$E(y_t|\mathcal{F}_{t-1}) = \mu^2 + \sigma^2 h_{t-1}^2, \quad \mathrm{var}(y_t|\mathcal{F}_{t-1}) = 2\left(\sigma^4 h_{t-1}^4 + 2\,\sigma^2\mu^2\, h_{t-1}^2\right),$$

where $\mathcal{F}_{t-1}$ expresses the sigma-algebra generated by the $\{\epsilon_s\,,\ s \le t-1\}$. The (time domain) log likelihood, under the pretence that the $y_t$ are Gaussian, is

$$\begin{aligned}
&Q_T^{time}(\sigma^2,\, w,\, \theta) = -\frac{T}{2}\,\ln(4\pi) - T\,\ln(\sigma^2)\\
&-\frac{1}{2}\,\ln\left(\prod_{t=1}^{T} 2(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta))\right)\\
&-\frac{1}{4}\sum_{t=1}^{T}(\frac{y_t}{\sigma^2} - w - h_{t-1}^2(\theta))^2/(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta)),
\end{aligned}$$

where $w = \mu^2/\sigma^2$. Expression (14) follows readily by applying the LM principle to $Q_T^{time}(\cdot,\cdot,\cdot)$. For an integer $m$ with $m < T$ set

$$\phi_m = \frac{1}{(T\,e(w)\,\sigma^2)^{\frac{1}{2}}}\Gamma_m^{\frac{1}{2}}\sum_{i=1}^{m}\tau_i\,\eta_i,$$

obtained by substituting $T-1$ with $m$ and $\hat{\sigma}^2$, $\hat{w}$ with $\sigma^2$, $w$ in $\phi_{T-1}(\hat{\sigma}^2, \hat{w})$ (cf. (15)) and writing $\eta_i = \eta_i(\sigma^2, w)$.

**Proof of Theorem 4.** This follows adapting the proof of Robinson (1991, Theorem 2.1). Setting $\bar{\phi}_m = (T\,e(w)\,\sigma^2)^{-\frac{1}{2}}\Gamma^{-\frac{1}{2}}\sum_{i=1}^{m}\tau_i\bar{\eta}_i$, $\bar{\eta}_i = \sum_{t=1}^{T}\epsilon_{t-i}\,X_t\,,$ from $E(X_t\epsilon_{t-i}|\mathcal{F}_{t-1}) = 0$, for any $i \ge 1$ it follows that $E\,\|\,\sum_{i=1}^{m}\tau_i\tilde{\eta}_i\,\|^2 = O(m^2)$ with $\tilde{\eta}_i = \bar{\eta}_i - \eta_i = \sum_{t=1}^{i}\epsilon_{t-i}\,X_t\,,$ yielding $\phi_m - \bar{\phi}_m = o_p(1)$. Under $H_0$ the covariance matrix of the score is block-diagonal with respect to $(\sigma^2, w)$ on one hand and $\theta$ on the other, yielding the usual expression for the LM test statistic (14). Note that by square summability of the $\tau_i$, $\lim_{t\to\infty}\sum_{i=t}^{\infty}\tau_i\tau_i' = 0$, the objective function $Q_T^{time}(\cdot,\cdot,\cdot)$ and the conditional objective function, obtained substituting the $\epsilon_t$ with the $\bar{\epsilon}_t = \epsilon_t 1_{t>0}$, yield asymptotically equivalent scores. Further, from simple but tedious calculations, under $H_0$

$$E\left(\frac{\partial Q_T^{time}(\sigma^2, w, 0)}{\partial\theta}\frac{\partial Q_T^{time}(\sigma^2, w, 0)}{\partial\theta'}\right) = \frac{e(w)}{(1+2w)^2}\,\sigma^2\Gamma_{T-1}\,,$$

and $E(X_t^2 \mid \mathcal{F}_{t-1}) = e(w)$. Under Assumptions $C$, in particular $C_2$ and $D$, a martingale CLT (see Brown (1971)) applies, yielding, as $T \to \infty$ and finite $m$,

$$\bar{\phi}_m \to_d \mathcal{N}_p(0, I_p).$$

Under $H_0$, $y_t^2 - E(y_t^2)$ and $y_t - E(y_t)$ are stationary square integrable martingale differences and thus $\hat{\sigma}^2 \to_p \sigma^2$ and $\hat{w} \to_p w$, as $T \to \infty$ by Slutzky lemma ($\to_p$ denoting convergence in probability ), yielding

$$\eta_i(\hat{\sigma}^2, \hat{w}) - \eta_i = \sum_{t=i+1}^{T} \big((x_{t-i} - \hat{\mu})X_t(\hat{\sigma}^2, \hat{w}) - (x_{t-i} - \mu)X_t\big) = o_p(T^{\frac{1}{2}}) \,.$$

Finally from $E(\bar{\eta}_i \, \bar{\eta}_{i+j}) = e(w) \, T \, \sigma^2 \delta(0, j)$, $i \neq 0$ we get as $m \to \infty$

$$E \parallel \sum_{i=m}^{T-1} \tau_i \bar{\eta}_i \parallel^2 \leq T e(w) \, \sigma^2 \sum_{i=m}^{\infty} \parallel \tau_i \parallel^2 = o(T)$$

and Bernstein's Lemma (Hannan 1970, p.242) applies. $\square$

## D.1 The robustified LM test

We relax Assumption $D$ as follows:
**Assumptions D'.**

$D_1'$ $E(\epsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$, $E[(\epsilon_t^2 - \sigma^2)\epsilon_t^2|\mathcal{F}_{t-1}] = 2\sigma^4$ .
$D_2'$ $E(\epsilon_1^{16}) < \infty$ .

$D_3'$ *For* $k, r \geq 1$

$$E(X_1 X_{1+i} \epsilon_{1-k} \epsilon_{1-r}^a) = \begin{cases} 0\,, & a = 0\,, \\ \begin{cases} 0\,, & i \geq 1, \\ f_k\,, & i = 0,\, k = r\,, \end{cases} & a = 1. \end{cases}$$

$D_4'$ $\inf_{i \geq 1} \;\; E(X_1^2 \epsilon_{1-i}^2) > 0$ .

$D_5'$ $\frac{1}{T} \sum_{t=1}^{T} \epsilon_{t-i} \epsilon_{t-j} E(X_t^2|\mathcal{F}_{t-1}) \to_{a.s.} f_i \delta(i, j)$, *as* $T \to \infty$.

Assumptions $D'$ allow a great deal of heterogeneity in the $\epsilon_t$. In fact we only need the $\epsilon_t$ to have constant conditional moment up to the fourth order, the trade off being the strict unconditional moment condition $D_2'$. Assumption $D_3'$ expresses the minimal degree of stationarity required in the $\epsilon_t$ and Assumption $D_4'$ guarantees that the distribution of the score is non-singular asymptotically. Assumption $D_5'$ is a mild ergodicity condition.

The robustified test statistic is

$$\Phi_R = \phi_R' \phi_R, \tag{46}$$

with $\phi_R = (T\hat{\Gamma}_R)^{-\frac{1}{2}} \sum_{i=1}^{T-1} \tau_i \eta_i(\hat{\sigma^2}, \hat{w})$ setting $\hat{\Gamma}_R = \sum_{i=1}^{T-1} \tau_i \tau_i' \bar{f}_i(\sigma^2, w)$, $\bar{f}_i(\sigma^2, w) = \frac{1}{T} \sum_{t=i+1}^{T} X_t^2(\sigma^2, w)(x_{t-i} - \bar{x})^2$. A frequency domain expression can be readily obtained. The considerations made for $\Phi$ concerning the computational gains are even more important for the $\Phi_R$ statistic, owing to the $\bar{f}_i(\sigma^2, w)$. We now establish the asymptotic distribution of $\Phi_R$ under $H_0$.

**Theorem 7** *Under the null hypothesis $H_0$ and Assumptions $C$ and $D'$, as $T \to \infty$,*

$$\Phi_R \to_d \chi_p^2 .$$

**Proof.** The proof closely follows that of Theorem 4 and therefore it will be sketched only. The block-diagonality of the variance-covariance matrix of the score and the substitution of the $\bar{\eta}_i$ with $\eta_i$ follow from Assumptions $D_1', D_2'$. Then by Assumptions $D_1', D_3', D_5'$ we obtain that, as $T \to \infty$,

$$\frac{1}{T^{\frac{1}{2}}} \sum_{i=1}^{m} \tau_i \bar{\eta}_i \to_d \mathcal{N}_p(0, \sum_{i=1}^{m} \tau_i \tau_i f_i) .$$

Assumption $D_4'$ guarantees that the matrix $\Gamma_R$ is invertible. Assumption $D_1'$ yields $X_t - X_t(\hat{\sigma^2}, \hat{w}) = o_p(1)$ and thus, for fixed $m$ and $T \to \infty$, $\sum_{i=1}^{m} \tau_i \tau_i'(\bar{f}_i - \bar{f}_i(\hat{\sigma^2}, \hat{w})) \to_p 0$, and $\sum_{i=1}^{m} \tau_i \tau_i'(\bar{f}_i - f_i) \to_p 0$ by $D_5'$ and $E|\bar{f}_i| = O(1)$, $E|\bar{f}_i(\hat{\sigma^2}, \hat{w})| = O(1)$ by $D_2'$. Thus, as $m$ (and then $T$) goes to infinity,

$$\sum_{i=m}^{T-1} \tau_i \tau_i'(E|\bar{f}_i| + E|\bar{f}_i(\hat{\sigma^2}, \hat{w})| + f_i) \to_p 0.\square$$

# E A feasible parameterization: the $ARFIMA(p, d, q)$ coefficients

In this section we show that the parameterization (9) employed in the empirical application of section 6 satisfies Assumptions $B$. Let us start by considering the case $p = q = 0$, $d(\theta) = \theta$, setting

$$c_i(\theta) = \begin{cases} \prod_{k=1}^{i} \frac{k-1+\theta}{k} & , i \geq 1 , \\ 1 & , i = 0 , \end{cases} \tag{47}$$

and $\alpha_i(\theta) = c_i(\theta)$, with $\alpha(\lambda) = (1 - e^{i\lambda})^{-\theta}$ $(-\pi \leq \lambda < \pi)$. Using Stirling's formula (Brockwell and Davis 1987, p.522), $\alpha_i(\theta) \sim k\, i^{\theta-1}$ , as $i \to \infty$, and

$B_1$ holds. From (47) one gets $\alpha_i(\theta) - \alpha_{i+1}(\theta) = \alpha_i(\theta)\left(\frac{1-\theta}{i+1}\right)$ and thus $B_2$ is trivially satisfied. Concerning $B_3$, $\ln\, \alpha_i(\theta) = \sum_{k=1}^{i} \ln\left((k-1+\theta)/k\right)$. Differentiating yields $\frac{\partial}{\partial\theta} \ln\, \alpha_i(\theta) = \sum_{k=1}^{i} \ln\left(\frac{1}{k-1+\theta}\right)$. As $n \to \infty$, $\sum_{k=1}^{n} \frac{1}{k} \sim \ln(n)$, yielding $\frac{\partial}{\partial\theta}\alpha_i(\theta) \sim k\, i^{\theta-1} \ln(i)$ as $i \to \infty$. Likewise from $\frac{\partial}{\partial\theta} \ln\, \alpha_i(\theta) - \frac{\partial}{\partial\theta} \ln\, \alpha_{i+1}(\theta) = \ln(i+\theta)$, as $i \to \infty$,

$$\frac{\partial}{\partial\theta}\alpha_i(\theta) - \frac{\partial}{\partial\theta}\alpha_{i+1}(\theta) = \frac{1-\theta}{i+\theta}\frac{\partial}{\partial\theta}\alpha_{i+1}(\theta) + \alpha_i(\theta)\ln(i+\theta)\ \sim\ k\alpha_i(\theta)\ \ln(i),$$

and thus $B_4$ holds. By an identical argument it can be shown that Assumptions $B_5$ and $B_6$ are satisfied. Finally, regarding $B_7$, imposing the condition $\alpha_i(\theta) = \alpha_i(\theta^*)$, for any integer $i \geq 1$ yields

$$(\theta^i - (\theta^*)^i)c_i + (\theta^{i-1} - (\theta^*)^{i-1})c_{i-1} + \ldots + (\theta - \theta^*)c_1 = 0,$$

where the coefficients $c_j\,(j = 1, \ldots, i)$ are positive functions of the integers $1, 2, \ldots, i$ only. Hence the above condition is true if and only if $\theta = \theta^*$. $B_8$ follows stacking $\partial\bar{\gamma}_{yy}(u;\psi)/\partial\psi$, for $u = 0, 1, 2$, in a $3 \times 3$ matrix and showing, by algebraic calculations, that it is nonsingular. These arguments can be easily extended to the case $p, q > 0$, setting:

$$\alpha_i(\theta) = \phi_1\alpha_{i-1}(\theta) + .. + \phi_p\alpha_{i-p}(\theta) + c_i(d) + \chi_1 c_{i-1}(d) + .. + \chi_q c_{i-q}(d),$$

where $\theta = (\chi_1, .., \chi_q, \phi_1, .., \phi_p, d)'$.

Figure 1: **Correlograms for stock index returns series, levels and squares (first and second row), and foreign exchange rate returns, levels and squares (third and fourth row).**

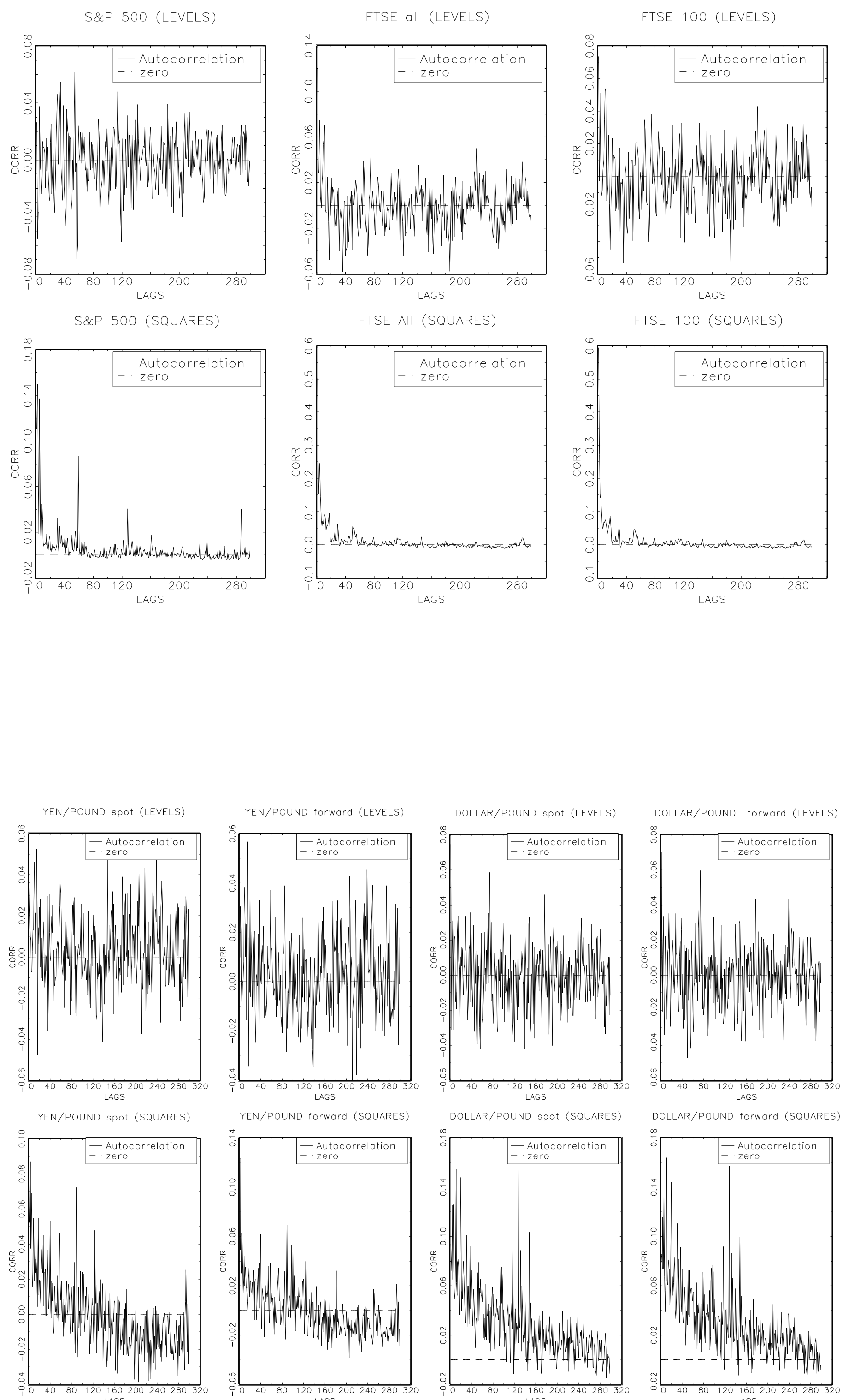

**Table 1.1**

Sample size: 1024

| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{\kappa}$ | 0 | -0.0888 | 0.0216 | $\bar{\kappa}$ | 0 | -0.2370 | 0.0211 | $\bar{\kappa}$ | 0 | -0.0828 | 0.0242 |
| $\phi_1$ | -0.8 | -0.7609 | 0.0030 | $\phi_1$ | -0.8 | -0.7530 | 0.0030 | $\phi_1$ | -0.8 | -0.7621 | 0.0039 |
| $d$ | 0.15 | 0.1430 | 0.0043 | $d$ | 0.25 | 0.1767 | 0.0045 | $d$ | 0.45 | 0.3411 | 0.0045 |
| $\bar{\kappa}$ | 0 | -0.1695 | 0.0221 | $\bar{\kappa}$ | 0 | -0.1176 | 0.0247 | $\bar{\kappa}$ | 0 | 0.0466 | 0.0258 |
| $\phi_1$ | -0.4 | -0.3323 | 0.0065 | $\phi_1$ | -0.4 | -0.3531 | 0.0064 | $\phi_1$ | -0.4 | -0.2604 | 0.0097 |
| $d$ | 0.15 | 0.0850 | 0.0023 | $d$ | 0.25 | 0.2017 | 0.0031 | $d$ | 0.45 | 0.4041 | 0.0030 |
| $\bar{\kappa}$ | 0 | -0.1752 | 0.0257 | $\bar{\kappa}$ | 0 | -0.1651 | 0.0247 | $\bar{\kappa}$ | 0 | 0.2292 | 0.0237 |
| $\phi_1$ | 0.4 | 0.3475 | 0.0100 | $\phi_1$ | 0.4 | 0.3640 | 0.0106 | $\phi_1$ | 0.4 | 0.3782 | 0.0123 |
| $d$ | 0.15 | 0.0867 | 0.0027 | $d$ | 0.25 | 0.1556 | 0.0045 | $d$ | 0.45 | 0.3066 | 0.0050 |
| $\bar{\kappa}$ | 0 | 0.0975 | 0.0242 | $\bar{\kappa}$ | 0 | 0.0764 | 0.0232 | $\bar{\kappa}$ | 0 | 0.1831 | 0.0212 |
| $\phi_1$ | 0.8 | 0.7427 | 0.0054 | $\phi_1$ | 0.8 | 0.7606 | 0.0060 | $\phi_1$ | 0.8 | 0.7742 | 0.0073 |
| $d$ | 0.15 | 0.1203 | 0.0038 | $d$ | 0.25 | 0.1711 | 0.0045 | $d$ | 0.45 | 0.33906 | 0.0047 |
| $\bar{\kappa}$ | 1.5 | 1.5800 | 0.0586 | $\bar{\kappa}$ | 1.5 | 1.0974 | 0.0568 | $\bar{\kappa}$ | 1.5 | 1.5199 | 0.0648 |
| $\phi_1$ | -0.8 | -0.7611 | 0.0042 | $\phi_1$ | -0.8 | -0.7650 | 0.0041 | $\phi_1$ | -0.8 | -0.7435 | 0.0063 |
| $d$ | 0.15 | 0.1257 | 0.0046 | $d$ | 0.25 | 0.1536 | 0.0048 | $d$ | 0.45 | 0.3048 | 0.0054 |
| $\bar{\kappa}$ | 1.5 | 1.3448 | 0.0603 | $\bar{\kappa}$ | 1.5 | 0.8106 | 0.0596 | $\bar{\kappa}$ | 1.5 | 1.3842 | 0.0638 |
| $\phi_1$ | -0.4 | -0.3376 | 0.0067 | $\phi_1$ | -0.4 | -0.3440 | 0.0070 | $\phi_1$ | -0.4 | -0.2270 | 0.00110 |
| $d$ | 0.15 | 0.0769 | 0.0025 | $d$ | 0.25 | 0.1826 | 0.0033 | $d$ | 0.45 | 0.3862 | 0.0037 |
| $\bar{\kappa}$ | 1.5 | 1.0881 | 0.0679 | $\bar{\kappa}$ | 1.5 | 0.9772 | 0.0635 | $\bar{\kappa}$ | 1.5 | 1.5241 | 0.0638 |
| $\phi_1$ | 0.4 | 0.3505 | 0.0111 | $\phi_1$ | 0.4 | 0.4013 | 0.0091 | $\phi_1$ | 0.4 | 0.4078 | 0.0120 |
| $d$ | 0.15 | 0.0761 | 0.0027 | $d$ | 0.25 | 0.1491 | 0.0041 | $d$ | 0.45 | 0.2950 | 0.0053 |
| $\bar{\kappa}$ | 1.5 | 1.4940 | 0.0634 | $\bar{\kappa}$ | 1.5 | 1.4547 | 0.0585 | $\bar{\kappa}$ | 1.5 | 1.6754 | 0.0525 |
| $\phi_1$ | 0.8 | 0.7155 | 0.0068 | $\phi_1$ | 0.8 | 0.7353 | 0.0074 | $\phi_1$ | 0.8 | 0.7576 | 0.0085 |
| $d$ | 0.15 | 0.1232 | 0.0036 | $d$ | 0.25 | 0.1723 | 0.0042 | $d$ | 0.45 | 0.3470 | 0.0046 |
| $\bar{\kappa}$ | 3.0 | 2.8591 | 0.0951 | $\bar{\kappa}$ | 3.0 | 2.4178 | 0.0975 | $\bar{\kappa}$ | 3.0 | 2.7228 | 0.1042 |
| $\phi_1$ | -0.8 | -0.7579 | 0.0046 | $\phi_1$ | -0.8 | -0.7514 | 0.0047 | $\phi_1$ | -0.8 | -0.7341 | 0.0067 |
| $d$ | 0.15 | 0.1166 | 0.0044 | $d$ | 0.25 | 0.1489 | 0.0050 | $d$ | 0.45 | 0.2951 | 0.0056 |
| $\bar{\kappa}$ | 3.0 | 2.8950 | 0.0927 | $\bar{\kappa}$ | 3.0 | 2.3159 | 0.0927 | $\bar{\kappa}$ | 3.0 | 2.2586 | 0.0987 |
| $\phi_1$ | -0.4 | -0.3571 | 0.0069 | $\phi_1$ | -0.4 | -0.3671 | 0.0081 | $\phi_1$ | -0.4 | -0.2525 | 0.0108 |
| $d$ | 0.15 | 0.0768 | 0.0026 | $d$ | 0.25 | 0.1819 | 0.0039 | $d$ | 0.45 | 0.3764 | 0.0040 |
| $\bar{\kappa}$ | 3.0 | 1.9126 | 0.1070 | $\bar{\kappa}$ | 3.0 | 2.0682 | 0.1054 | $\bar{\kappa}$ | 3.0 | 2.4803 | 0.1031 |
| $\phi_1$ | 0.4 | 0.3755 | 0.0091 | $\phi_1$ | 0.4 | 0.4036 | 0.0095 | $\phi_1$ | 0.4 | 0.4052 | 0.0123 |
| $d$ | 0.15 | 0.0797 | 0.0032 | $d$ | 0.25 | 0.1406 | 0.0043 | $d$ | 0.45 | 0.2830 | 0.0053 |
| $\bar{\kappa}$ | 3.0 | 2.4853 | 0.0975 | $\bar{\kappa}$ | 3.0 | 2.5792 | 0.0942 | $\bar{\kappa}$ | 3.0 | 2.4803 | 0.1031 |
| $\phi_1$ | 0.8 | 0.7212 | 0.0067 | $\phi_1$ | 0.8 | 0.7471 | 0.0067 | $\phi_1$ | 0.8 | 0.7642 | 0.0075 |
| $d$ | 0.15 | 0.1191 | 0.0033 | $d$ | 0.25 | 0.1748 | 0.0043 | $d$ | 0.45 | 0.3592 | 0.0044 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 1.2**

Sample size: 2048

| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{\kappa}$ | 0 | -0.0631 | 0.0197 | $\bar{\kappa}$ | 0 | -0.1923 | 0.0196 | $\bar{\kappa}$ | 0 | -0.0544 | 0.0214 |
| $\phi_1$ | -0.8 | -0.7773 | 0.0021 | $\phi_1$ | -0.8 | -0.7712 | 0.0019 | $\phi_1$ | -0.8 | -0.7834 | 0.0041 |
| $d$ | 0.15 | 0.1480 | 0.0038 | $d$ | 0.25 | 0.1856 | 0.0041 | $d$ | 0.45 | 0.3814 | 0.0041 |
| $\bar{\kappa}$ | 0 | -0.1467 | 0.0207 | $\bar{\kappa}$ | 0 | -0.1257 | 0.0251 | $\bar{\kappa}$ | 0 | -0.0665 | 0.0253 |
| $\phi_1$ | -0.4 | -0.3445 | 0.0056 | $\phi_1$ | -0.4 | -0.3481 | 0.0051 | $\phi_1$ | -0.4 | -0.3109 | 0.0077 |
| $d$ | 0.15 | 0.0943 | 0.0037 | $d$ | 0.25 | 0.2186 | 0.0021 | $d$ | 0.45 | 0.4161 | 0.0025 |
| $\bar{\kappa}$ | 0 | -0.2168 | 0.0246 | $\bar{\kappa}$ | 0 | -0.1452 | 0.0227 | $\bar{\kappa}$ | 0 | 0.1079 | 0.0225 |
| $\phi_1$ | 0.4 | 0.3695 | 0.0078 | $\phi_1$ | 0.4 | 0.3797 | 0.0073 | $\phi_1$ | 0.4 | 0.3706 | 0.0110 |
| $d$ | 0.15 | 0.1086 | 0.0023 | $d$ | 0.25 | 0.1933 | 0.0029 | $d$ | 0.45 | 0.3475 | 0.0041 |
| $\bar{\kappa}$ | 0 | 0.0902 | 0.0218 | $\bar{\kappa}$ | 0 | 0.0425 | 0.0201 | $\bar{\kappa}$ | 0 | 0.1061 | 0.0192 |
| $\phi_1$ | 0.8 | 0.7663 | 0.0043 | $\phi_1$ | 0.8 | 0.7737 | 0.0053 | $\phi_1$ | 0.8 | 0.7814 | 0.0062 |
| $d$ | 0.15 | 0.1227 | 0.0031 | $d$ | 0.25 | 0.1774 | 0.0038 | $d$ | 0.45 | 0.3591 | 0.0040 |
| $\bar{\kappa}$ | 1.5 | 1.3265 | 0.0552 | $\bar{\kappa}$ | 1.5 | 0.9625 | 0.0531 | $\bar{\kappa}$ | 1.5 | 1.3491 | 0.0574 |
| $\phi_1$ | -0.8 | -0.8069 | 0.0017 | $\phi_1$ | -0.8 | -0.8004 | 0.0012 | $\phi_1$ | -0.8 | -0.7812 | 0.0046 |
| $d$ | 0.15 | 0.1227 | 0.0042 | $d$ | 0.25 | 0.1834 | 0.0048 | $d$ | 0.45 | 0.3364 | 0.0052 |
| $\bar{\kappa}$ | 1.5 | 0.5476 | 0.0572 | $\bar{\kappa}$ | 1.5 | 0.9651 | 0.0529 | $\bar{\kappa}$ | 1.5 | 0.9345 | 0.0605 |
| $\phi_1$ | -0.4 | -0.3693 | 0.0058 | $\phi_1$ | -0.4 | -0.4258 | 0.0064 | $\phi_1$ | -0.4 | -0.2744 | 0.0091 |
| $d$ | 0.15 | 0.0931 | 0.0013 | $d$ | 0.25 | 0.2421 | 0.0028 | $d$ | 0.45 | 0.3959 | 0.0032 |
| $\bar{\kappa}$ | 1.5 | 1.1763 | 0.0585 | $\bar{\kappa}$ | 1.5 | 1.2395 | 0.0609 | $\bar{\kappa}$ | 1.5 | 1.5391 | 0.0562 |
| $\phi_1$ | 0.4 | 0.4382 | 0.0033 | $\phi_1$ | 0.4 | 0.3733 | 0.0038 | $\phi_1$ | 0.4 | 0.4323 | 0.0109 |
| $d$ | 0.15 | 0.0889 | 0.0019 | $d$ | 0.25 | 0.1929 | 0.0037 | $d$ | 0.45 | 0.3321 | 0.0046 |
| $\bar{\kappa}$ | 1.5 | 1.8294 | 0.0634 | $\bar{\kappa}$ | 1.5 | 1.8599 | 0.0543 | $\bar{\kappa}$ | 1.5 | 1.5616 | 0.0467 |
| $\phi_1$ | 0.8 | 0.7208 | 0.0044 | $\phi_1$ | 0.8 | 0.7485 | 0.0052 | $\phi_1$ | 0.8 | 0.7514 | 0.0079 |
| $d$ | 0.15 | 0.1442 | 0.0023 | $d$ | 0.25 | 0.2197 | 0.0035 | $d$ | 0.45 | 0.3547 | 0.0044 |
| $\bar{\kappa}$ | 3.0 | 2.4782 | 0.0912 | $\bar{\kappa}$ | 3.0 | 2.3934 | 0.0819 | $\bar{\kappa}$ | 3.0 | 4.4638 | 0.0891 |
| $\phi_1$ | -0.8 | -0.7718 | 0.0038 | $\phi_1$ | -0.8 | -0.7962 | 0.0034 | $\phi_1$ | -0.8 | -0.8681 | 0.0017 |
| $d$ | 0.15 | 0.0779 | 0.0032 | $d$ | 0.25 | 0.2183 | 0.0041 | $d$ | 0.45 | 0.3928 | 0.0054 |
| $\bar{\kappa}$ | 3.0 | 2.8291 | 0.0854 | $\bar{\kappa}$ | 3.0 | 2.3934 | 0.0819 | $\bar{\kappa}$ | 3.0 | 3.7839 | 0.0854 |
| $\phi_1$ | -0.4 | -0.3908 | 0.0058 | $\phi_1$ | -0.4 | -0.3735 | 0.0071 | $\phi_1$ | -0.4 | -0.1072 | 0.0085 |
| $d$ | 0.15 | 0.0955 | 0.0024 | $d$ | 0.25 | 0.2288 | 0.0032 | $d$ | 0.45 | 0.4651 | 0.0018 |
| $\bar{\kappa}$ | 3.0 | 1.9903 | 0.1037 | $\bar{\kappa}$ | 3.0 | 2.3729 | 0.0888 | $\bar{\kappa}$ | 3.0 | 3.5272 | 0.0789 |
| $\phi_1$ | 0.4 | 0.3965 | 0.0079 | $\phi_1$ | 0.4 | 0.2303 | 0.0060 | $\phi_1$ | 0.4 | 0.6246 | 0.0043 |
| $d$ | 0.15 | 0.0903 | 0.0028 | $d$ | 0.25 | 0.2844 | 0.0051 | $d$ | 0.45 | 0.4095 | 0.0037 |
| $\bar{\kappa}$ | 3.0 | 2.3209 | 0.0819 | $\bar{\kappa}$ | 3.0 | 3.3355 | 0.0839 | $\bar{\kappa}$ | 3.0 | 2.7774 | 0.0686 |
| $\phi_1$ | 0.8 | 0.7189 | 0.0061 | $\phi_1$ | 0.8 | 0.6375 | 0.0176 | $\phi_1$ | 0.8 | 0.8411 | 0.0078 |
| $d$ | 0.15 | 0.1236 | 0.0029 | $d$ | 0.25 | 0.2702 | 0.0041 | $d$ | 0.45 | 0.4423 | 0.0024 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 1.3**

Sample size: 4096

| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{\kappa}$ | 0 | -0.0119 | 0.0170 | $\bar{\kappa}$ | 0 | -0.1248 | 0.0176 | $\bar{\kappa}$ | 0 | 0.0633 | 0.0183 |
| $\phi_1$ | -0.8 | -0.7879 | 0.0013 | $\phi_1$ | -0.8 | -0.7841 | 0.0013 | $\phi_1$ | -0.8 | -0.7942 | 0.0017 |
| $d$ | 0.15 | 0.1382 | 0.0032 | $d$ | 0.25 | 0.1900 | 0.0037 | $d$ | 0.45 | 0.4113 | 0.0031 |
| $\bar{\kappa}$ | 0 | -0.1712 | 0.0185 | $\bar{\kappa}$ | 0 | -0.0409 | 0.0242 | $\bar{\kappa}$ | 0 | -0.0547 | 0.0235 |
| $\phi_1$ | -0.4 | -0.3747 | 0.0036 | $\phi_1$ | -0.4 | -0.3783 | 0.0036 | $\phi_1$ | -0.4 | -0.3017 | 0.0074 |
| $d$ | 0.15 | 0.1081 | 0.0017 | $d$ | 0.25 | 0.2376 | 0.0016 | $d$ | 0.45 | 0.4252 | 0.0023 |
| $\bar{\kappa}$ | 0 | -0.1888 | 0.0231 | $\bar{\kappa}$ | 0 | -0.2074 | 0.0202 | $\bar{\kappa}$ | 0 | 0.2564 | 0.0166 |
| $\phi_1$ | 0.4 | 0.3925 | 0.0046 | $\phi_1$ | 0.4 | 0.4152 | 0.0052 | $\phi_1$ | 0.4 | 0.4298 | 0.0039 |
| $d$ | 0.15 | 0.1223 | 0.0017 | $d$ | 0.25 | 0.2061 | 0.0023 | $d$ | 0.45 | 0.4320 | 0.0024 |
| $\bar{\kappa}$ | 0 | 0.0463 | 0.0187 | $\bar{\kappa}$ | 0 | 0.1781 | 0.0155 | $\bar{\kappa}$ | 0 | 0.2454 | 0.0163 |
| $\phi_1$ | 0.8 | 0.7771 | 0.0037 | $\phi_1$ | 0.8 | 0.8027 | 0.0029 | $\phi_1$ | 0.8 | 0.8335 | 0.0039 |
| $d$ | 0.15 | 0.1267 | 0.0027 | $d$ | 0.25 | 0.2192 | 0.0039 | $d$ | 0.45 | 0.4292 | 0.0023 |
| $\bar{\kappa}$ | 1.5 | 1.6228 | 0.0484 | $\bar{\kappa}$ | 1.5 | 1.3077 | 0.0472 | $\bar{\kappa}$ | 1.5 | 1.5333 | 0.0512 |
| $\phi_1$ | -0.8 | -0.7950 | 0.0022 | $\phi_1$ | -0.8 | -0.7930 | 0.0021 | $\phi_1$ | -0.8 | -0.7832 | 0.0022 |
| $d$ | 0.15 | 0.1073 | 0.0035 | $d$ | 0.25 | 0.1546 | 0.0042 | $d$ | 0.45 | 0.3712 | 0.0044 |
| $\bar{\kappa}$ | 1.5 | 1.3475 | 0.0484 | $\bar{\kappa}$ | 1.5 | 1.2432 | 0.0501 | $\bar{\kappa}$ | 1.5 | 1.5295 | 0.0531 |
| $\phi_1$ | -0.4 | -0.3980 | 0.0044 | $\phi_1$ | -0.4 | -0.3839 | 0.0058 | $\phi_1$ | -0.4 | -0.3390 | 0.0065 |
| $d$ | 0.15 | 0.1089 | 0.0021 | $d$ | 0.25 | 0.2433 | 0.0029 | $d$ | 0.45 | 0.3974 | 0.0029 |
| $\bar{\kappa}$ | 1.5 | 0.8611 | 0.0587 | $\bar{\kappa}$ | 1.5 | 1.9787 | 0.0564 | $\bar{\kappa}$ | 1.5 | 1.0910 | 0.0559 |
| $\phi_1$ | 0.4 | 0.4062 | 0.0059 | $\phi_1$ | 0.4 | 0.4176 | 0.0062 | $\phi_1$ | 0.4 | 0.3914 | 0.0103 |
| $d$ | 0.15 | 0.1174 | 0.0027 | $d$ | 0.25 | 0.1916 | 0.0036 | $d$ | 0.45 | 0.3624 | 0.0035 |
| $\bar{\kappa}$ | 1.5 | 1.3712 | 0.0487 | $\bar{\kappa}$ | 1.5 | 1.4489 | 0.0454 | $\bar{\kappa}$ | 1.5 | 1.7064 | 0.0413 |
| $\phi_1$ | 0.8 | 0.7643 | 0.0044 | $\phi_1$ | 0.8 | 0.7748 | 0.0051 | $\phi_1$ | 0.8 | 0.7706 | 0.0067 |
| $d$ | 0.15 | 0.1173 | 0.0026 | $d$ | 0.25 | 0.1916 | 0.0036 | $d$ | 0.45 | 0.3998 | 0.0034 |
| $\bar{\kappa}$ | 3.0 | 2.9487 | 0.0810 | $\bar{\kappa}$ | 3.0 | 4.4445 | 0.0761 | $\bar{\kappa}$ | 3.0 | 4.7166 | 0.0843 |
| $\phi_1$ | -0.8 | -0.7924 | 0.0032 | $\phi_1$ | -0.8 | -0.8650 | 0.0010 | $\phi_1$ | -0.8 | -0.8729 | 0.0014 |
| $d$ | 0.15 | 0.0927 | 0.0034 | $d$ | 0.25 | 0.1898 | 0.0055 | $d$ | 0.45 | 0.4008 | 0.0054 |
| $\bar{\kappa}$ | 3.0 | 2.8843 | 0.0755 | $\bar{\kappa}$ | 3.0 | 1.9387 | 0.0886 | $\bar{\kappa}$ | 3.0 | 2.5579 | 0.0751 |
| $\phi_1$ | -0.4 | -0.4077 | 0.0046 | $\phi_1$ | -0.4 | -0.3822 | 0.0065 | $\phi_1$ | -0.4 | -0.3088 | 0.0074 |
| $d$ | 0.15 | 0.1102 | 0.0025 | $d$ | 0.25 | 0.2420 | 0.0027 | $d$ | 0.45 | 0.3933 | 0.0034 |
| $\bar{\kappa}$ | 3.0 | 1.5057 | 0.0924 | $\bar{\kappa}$ | 3.0 | 1.9387 | 0.0886 | $\bar{\kappa}$ | 3.0 | 2.5579 | 0.0852 |
| $\phi_1$ | 0.4 | 0.4087 | 0.0062 | $\phi_1$ | 0.4 | 0.4131 | 0.0069 | $\phi_1$ | 0.4 | 0.3872 | 0.0099 |
| $d$ | 0.15 | 0.0935 | 0.0024 | $d$ | 0.25 | 0.1765 | 0.0032 | $d$ | 0.45 | 0.3522 | 0.0042 |
| $\bar{\kappa}$ | 3.0 | 2.5540 | 0.0774 | $\bar{\kappa}$ | 3.0 | 2.4468 | 0.0751 | $\bar{\kappa}$ | 3.0 | 2.9603 | 0.0722 |
| $\phi_1$ | 0.8 | 0.7910 | 0.0027 | $\phi_1$ | 0.8 | 0.7650 | 0.0057 | $\phi_1$ | 0.8 | 0.7877 | 0.0061 |
| $d$ | 0.15 | 0.1358 | 0.0041 | $d$ | 0.25 | 0.1903 | 0.0035 | $d$ | 0.45 | 0.4008 | 0.0033 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 2.1**

| Sample size: 1024 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
| $\bar{\kappa}$ | 0 | -0.2033 | 0.0226 | $\bar{\kappa}$ | 0 | -0.1382 | 0.0253 | $\bar{\kappa}$ | 0 | 0.0178 | 0.0265 |
| $\omega_1$ | 0.5 | 0.4982 | 0.0008 | $\omega_1$ | 0.5 | 0.4967 | 0.0009 | $\omega_1$ | 0.5 | 0.4933 | 0.0011 |
| $\phi_1$ | -0.4 | -0.3303 | 0.0062 | $\phi_1$ | -0.4 | -0.3367 | 0.0066 | $\phi_1$ | -0.4 | -0.2526 | 0.0108 |
| $d$ | 0.15 | 0.0876 | 0.0022 | $d$ | 0.25 | 0.1978 | 0.0030 | $d$ | 0.45 | 0.4013 | 0.0031 |
| $\bar{\kappa}$ | 0 | -0.1562 | 0.0252 | $\bar{\kappa}$ | 0 | -0.1577 | 0.0244 | $\bar{\kappa}$ | 0 | 0.1490 | 0.0238 |
| $\omega_1$ | 0.5 | 0.4921 | 0.0013 | $\omega_1$ | 0.5 | 0.4934 | 0.0014 | $\omega_1$ | 0.5 | 0.4901 | 0.0015 |
| $\phi_1$ | 0.4 | 0.3383 | 0.0101 | $\phi_1$ | 0.4 | 0.3756 | 0.0104 | $\phi_1$ | 0.4 | 0.3659 | 0.0127 |
| $d$ | 0.15 | 0.0942 | 0.0028 | $d$ | 0.25 | 0.1587 | 0.0037 | $d$ | 0.45 | 0.3132 | 0.0049 |
| $\bar{\kappa}$ | 1.5 | 1.2053 | 0.0685 | $\bar{\kappa}$ | 1.5 | 1.0502 | 0.0673 | $\bar{\kappa}$ | 1.5 | 1.1679 | 0.0649 |
| $\omega_1$ | 0.5 | 0.4978 | 0.0009 | $\omega_1$ | 0.5 | 0.4919 | 0.0014 | $\omega_1$ | 0.5 | 0.4966 | 0.0009 |
| $\phi_1$ | -0.4 | -0.3429 | 0.0065 | $\phi_1$ | -0.4 | -0.3546 | 0.0076 | $\phi_1$ | -0.4 | -0.2573 | 0.0111 |
| $d$ | 0.15 | 0.0810 | 0.0025 | $d$ | 0.25 | 0.1846 | 0.0076 | $d$ | 0.45 | 0.3804 | 0.0038 |
| $\bar{\kappa}$ | 1.5 | 1.0502 | 0.0673 | $\bar{\kappa}$ | 1.5 | 1.1679 | 0.0649 | $\bar{\kappa}$ | 1.5 | 1.7450 | 0.0612 |
| $\omega_1$ | 0.5 | 0.4919 | 0.0014 | $\omega_1$ | 0.5 | 0.4921 | 0.0015 | $\omega_1$ | 0.5 | 0.4911 | 0.0015 |
| $\phi_1$ | 0.4 | 0.3428 | 0.0109 | $\phi_1$ | 0.4 | 0.4102 | 0.0095 | $\phi_1$ | 0.4 | 0.4011 | 0.0128 |
| $d$ | 0.15 | 0.0812 | 0.0029 | $d$ | 0.25 | 0.1484 | 0.0043 | $d$ | 0.45 | 0.3076 | 0.0052 |
| $\bar{\kappa}$ | 3.0 | 2.9649 | 0.0923 | $\bar{\kappa}$ | 3.0 | 2.2467 | 0.0943 | $\bar{\kappa}$ | 3.0 | 2.2600 | 0.0999 |
| $\omega_1$ | 0.5 | 0.4974 | 0.0009 | $\omega_1$ | 0.5 | 0.4973 | 0.0009 | $\omega_1$ | 0.5 | 0.4933 | 0.0012 |
| $\phi_1$ | -0.4 | -0.3479 | 0.0066 | $\phi_1$ | -0.4 | -0.3573 | 0.0078 | $\phi_1$ | -0.4 | -0.2527 | 0.0110 |
| $d$ | 0.15 | 0.0787 | 0.0026 | $d$ | 0.25 | 0.1785 | 0.0038 | $d$ | 0.45 | 0.3715 | 0.0042 |
| $\bar{\kappa}$ | 3.0 | 2.068 | 0.1059 | $\bar{\kappa}$ | 3.0 | 2.3136 | 0.1038 | $\bar{\kappa}$ | 3.0 | 2.7811 | 0.1029 |
| $\omega_1$ | 0.5 | 0.4924 | 0.0014 | $\omega_1$ | 0.5 | 0.4909 | 0.0015 | $\omega_1$ | 0.5 | 0.4863 | 0.0015 |
| $\phi_1$ | 0.4 | 0.3717 | 0.0103 | $\phi_1$ | 0.4 | 0.4178 | 0.0093 | $\phi_1$ | 0.4 | 0.4084 | 0.0165 |
| $d$ | 0.15 | 0.0835 | 0.0031 | $d$ | 0.25 | 0.1507 | 0.0044 | $d$ | 0.45 | 0.32814 | 0.0054 |

**Parameterization:** $\beta(L;\zeta) = (1-\omega_1 L)^{-1}$, $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 2.2**

Sample size: 2048

| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{\kappa}$ | 0 | -0.1756 | 0.0209 | $\bar{\kappa}$ | 0 | -0.0888 | 0.0249 | $\bar{\kappa}$ | 0 | 0.0532 | 0.0250 |
| $\omega_1$ | 0.5 | 0.4992 | 0.0006 | $\omega_1$ | 0.5 | 0.4984 | 0.0006 | $\omega_1$ | 0.5 | 0.4966 | 0.0078 |
| $\phi_1$ | -0.4 | -0.3514 | 0.0051 | $\phi_1$ | -0.4 | -0.3602 | 0.0050 | $\phi_1$ | -0.4 | -0.2756 | 0.0088 |
| $d$ | 0.15 | 0.0981 | 0.0021 | $d$ | 0.25 | 0.2238 | 0.0021 | $d$ | 0.45 | 0.4172 | 0.0026 |
| $\bar{\kappa}$ | 0 | -0.1962 | 0.0238 | $\bar{\kappa}$ | 0 | -0.1733 | 0.0231 | $\bar{\kappa}$ | 0 | 0.1029 | 0.0221 |
| $\omega_1$ | 0.5 | 0.4973 | 0.0009 | $\omega_1$ | 0.5 | 0.4964 | 0.0010 | $\omega_1$ | 0.5 | 0.4937 | 0.0011 |
| $\phi_1$ | 0.4 | 0.3857 | 0.0069 | $\phi_1$ | 0.4 | 0.3916 | 0.0075 | $\phi_1$ | 0.4 | 0.3410 | 0.0173 |
| $d$ | 0.15 | 0.1056 | 0.0023 | $d$ | 0.25 | 0.1883 | 0.0029 | $d$ | 0.45 | 0.3485 | 0.0042 |
| $\bar{\kappa}$ | 1.5 | 1.3419 | 0.0534 | $\bar{\kappa}$ | 1.5 | 1.1364 | 0.0626 | $\bar{\kappa}$ | 1.5 | 0.9927 | 0.0594 |
| $\omega_1$ | 0.5 | 0.4979 | 0.0006 | $\omega_1$ | 0.5 | 0.4935 | 0.0005 | $\omega_1$ | 0.5 | 0.4976 | 0.0008 |
| $\phi_1$ | -0.4 | -0.3649 | 0.0058 | $\phi_1$ | -0.4 | -0.4146 | 0.0057 | $\phi_1$ | -0.4 | -0.4967 | 0.0008 |
| $d$ | 0.15 | 0.0929 | 0.0023 | $d$ | 0.25 | 0.2296 | 0.0028 | $d$ | 0.45 | 0.3940 | 0.0032 |
| $\bar{\kappa}$ | 1.5 | 1.8963 | 0.0627 | $\bar{\kappa}$ | 1.5 | 1.2007 | 0.0601 | $\bar{\kappa}$ | 1.5 | 0.4973 | 0.0603 |
| $\omega_1$ | 0.5 | 0.4986 | 0.0010 | $\omega_1$ | 0.5 | 0.4968 | 0.0011 | $\omega_1$ | 0.5 | 0.4917 | 0.0011 |
| $\phi_1$ | 0.4 | 0.3922 | 0.0079 | $\phi_1$ | 0.4 | 0.4229 | 0.0078 | $\phi_1$ | 0.4 | 0.3787 | 0.0112 |
| $d$ | 0.15 | 0.0895 | 0.0026 | $d$ | 0.25 | 0.1773 | 0.0036 | $d$ | 0.45 | 0.3195 | 0.0046 |
| $\bar{\kappa}$ | 3.0 | 2.907 | 0.0873 | $\bar{\kappa}$ | 3.0 | 2.3920 | 0.0817 | $\bar{\kappa}$ | 3.0 | 2.1439 | 0.0930 |
| $\omega_1$ | 0.5 | 0.4993 | 0.0007 | $\omega_1$ | 0.5 | 0.4983 | 0.0006 | $\omega_1$ | 0.5 | 0.4968 | 0.0009 |
| $\phi_1$ | -0.4 | -0.3842 | 0.0059 | $\phi_1$ | -0.4 | -0.3773 | 0.0069 | $\phi_1$ | -0.4 | -0.2871 | 0.0092 |
| $d$ | 0.15 | 0.0874 | 0.0024 | $d$ | 0.25 | 0.2217 | 0.0032 | $d$ | 0.45 | 0.3908 | 0.0036 |
| $\bar{\kappa}$ | 3.0 | 1.9337 | 0.0992 | $\bar{\kappa}$ | 3.0 | 2.1964 | 0.0982 | $\bar{\kappa}$ | 3.0 | 2.5204 | 0.0921 |
| $\omega_1$ | 0.5 | 0.4975 | 0.0010 | $\omega_1$ | 0.5 | 0.4959 | 0.0011 | $\omega_1$ | 0.5 | 0.4921 | 0.0012 |
| $\phi_1$ | 0.4 | 0.3942 | 0.0077 | $\phi_1$ | 0.4 | 0.4301 | 0.0078 | $\phi_1$ | 0.4 | 0.4069 | 0.0112 |
| $d$ | 0.15 | 0.0911 | 0.0027 | $d$ | 0.25 | 0.1650 | 0.0040 | $d$ | 0.45 | 0.3269 | 0.0048 |

**Parameterization:** $\beta(L;\zeta) = (1-\omega_1 L)^{-1}$, $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 2.3**

Sample size: 4096

| Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. | Parameter | Value | Mean | Std.Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{\kappa}$ | 0 | -0.1369 | 0.0201 | $\bar{\kappa}$ | 0 | -0.0866 | 0.0240 | $\bar{\kappa}$ | 0 | 0.0110 | 0.0230 |
| $\omega_1$ | 0.5 | 0.4998 | 0.0004 | $\omega_1$ | 0.5 | 0.4998 | 0.0004 | $\omega_1$ | 0.5 | 0.4983 | 0.0006 |
| $\phi_1$ | -0.4 | -0.3729 | 0.0034 | $\phi_1$ | -0.4 | -0.3708 | 0.0035 | $\phi_1$ | -0.4 | -0.3084 | 0.0067 |
| $d$ | 0.15 | 0.1100 | 0.0018 | $d$ | 0.25 | 0.2338 | 0.0016 | $d$ | 0.45 | 0.4275 | 0.0021 |
| | | | | | | | | | | | |
| $\bar{\kappa}$ | 0 | -0.1339 | 0.0231 | $\bar{\kappa}$ | 0 | -0.1312 | 0.0220 | $\bar{\kappa}$ | 0 | 0.0785 | 0.0196 |
| $\omega_1$ | 0.5 | 0.4975 | 0.0007 | $\omega_1$ | 0.5 | 0.4979 | 0.0007 | $\omega_1$ | 0.5 | 0.4986 | 0.0007 |
| $\phi_1$ | 0.4 | 0.4090 | 0.0048 | $\phi_1$ | 0.4 | 0.4048 | 0.0056 | $\phi_1$ | 0.4 | 0.3543 | 0.0100 |
| $d$ | 0.15 | 0.1197 | 0.0018 | $d$ | 0.25 | 0.2106 | 0.0023 | $d$ | 0.45 | 0.3865 | 0.0031 |
| | | | | | | | | | | | |
| $\bar{\kappa}$ | 1.5 | 1.3602 | 0.0479 | $\bar{\kappa}$ | 1.5 | 1.2867 | 0.0548 | $\bar{\kappa}$ | 1.5 | 1.0698 | 0.0577 |
| $\omega_1$ | 0.5 | 0.4992 | 0.0004 | $\omega_1$ | 0.5 | 0.4999 | 0.0005 | $\omega_1$ | 0.5 | 0.4967 | 0.0006 |
| $\phi_1$ | -0.4 | -0.3882 | 0.0043 | $\phi_1$ | -0.4 | -0.3820 | 0.0054 | $\phi_1$ | -0.4 | -0.2854 | 0.0057 |
| $d$ | 0.15 | 0.1091 | 0.0021 | $d$ | 0.25 | 0.2419 | 0.0023 | $d$ | 0.45 | 0.4037 | 0.0029 |
| | | | | | | | | | | | |
| $\bar{\kappa}$ | 1.5 | 0.9725 | 0.0586 | $\bar{\kappa}$ | 1.5 | 1.0288 | 0.0552 | $\bar{\kappa}$ | 1.5 | 1.6044 | 0.0510 |
| $\omega_1$ | 0.5 | 0.4979 | 0.0007 | $\omega_1$ | 0.5 | 0.4975 | 0.0008 | $\omega_1$ | 0.5 | 0.4933 | 0.0008 |
| $\phi_1$ | 0.4 | 0.4065 | 0.0068 | $\phi_1$ | 0.4 | 0.4163 | 0.0062 | $\phi_1$ | 0.4 | 0.3665 | 0.0106 |
| $d$ | 0.15 | 0.1037 | 0.0023 | $d$ | 0.25 | 0.1891 | 0.0031 | $d$ | 0.45 | 0.3676 | 0.0037 |
| | | | | | | | | | | | |
| $\bar{\kappa}$ | 3.0 | 2.9792 | 0.0783 | $\bar{\kappa}$ | 3.0 | 2.5281 | 0.0760 | $\bar{\kappa}$ | 3.0 | 2.0276 | 0.0873 |
| $\omega_1$ | 0.5 | 0.4997 | 0.0005 | $\omega_1$ | 0.5 | 0.5001 | 0.0004 | $\omega_1$ | 0.5 | 0.4978 | 0.0006 |
| $\phi_1$ | -0.4 | -0.3991 | 0.0050 | $\phi_1$ | -0.4 | -0.3721 | 0.0060 | $\phi_1$ | -0.4 | -0.3144 | 0.0069 |
| $d$ | 0.15 | 0.1066 | 0.0022 | $d$ | 0.25 | 0.2349 | 0.0025 | $d$ | 0.45 | 0.3976 | 0.0034 |
| | | | | | | | | | | | |
| $\bar{\kappa}$ | 3.0 | 1.7137 | 0.0931 | $\bar{\kappa}$ | 3.0 | 2.0497 | 0.0903 | $\bar{\kappa}$ | 3.0 | 2.8007 | 0.0823 |
| $\omega_1$ | 0.5 | 0.4978 | 0.0008 | $\omega_1$ | 0.5 | 0.4970 | 0.0008 | $\omega_1$ | 0.5 | 0.4965 | 0.0008 |
| $\phi_1$ | 0.4 | 0.4094 | 0.0067 | $\phi_1$ | 0.4 | 0.4086 | 0.0069 | $\phi_1$ | 0.4 | 0.3943 | 0.0104 |
| $d$ | 0.15 | 0.0954 | 0.0023 | $d$ | 0.25 | 0.1801 | 0.0034 | $d$ | 0.45 | 0.3673 | 0.0040 |

**Parameterization:** $\beta(L;\zeta) = (1-\omega_1 L)^{-1}$, $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 3.1**

Sample size: 1024

| Parameter | Value | Probability | Parameter | Value | Probability | Parameter | Value | Probability |
|---|---|---|---|---|---|---|---|---|
| | | | $\bar{\kappa} = 0$ | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.957 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.860 |
| | | | $\bar{\kappa} = 0$ | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.846 | $d$ | 0.25 | 0.897 | $d$ | 0.45 | 0.931 |
| | | | $\bar{\kappa} = 0$ | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.987 | $d$ | 0.25 | 0.770 | $d$ | 0.45 | 0.834 |
| | | | $\bar{\kappa} = 0$ | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.960 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.856 |
| | | | $\bar{\kappa} = 1.5$ | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.936 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.879 |
| | | | $\bar{\kappa} = 1.5$ | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.989 | $d$ | 0.25 | 0.866 | $d$ | 0.45 | 0.895 |
| | | | $\bar{\kappa} = 1.5$ | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.982 | $d$ | 0.25 | 0.970 | $d$ | 0.45 | 0.824 |
| | | | $\bar{\kappa} = 1.5$ | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.959 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.936 |
| | | | $\bar{\kappa} = 3$ | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.943 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.739 |
| | | | $\bar{\kappa} = 3$ | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.980 | $d$ | 0.25 | 0.897 | $d$ | 0.45 | 0.858 |
| | | | $\bar{\kappa} = 3$ | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.972 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.773 |
| | | | $\bar{\kappa} = 3$ | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.966 | $d$ | 0.25 | 1 | $d$ | 0.45 | 0.783 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$**. Design:** $\rho^2 = 1,\ \sigma^2 = 1,\ \mu = 0$.

**Theoretical probability: 0.95.**

**Table 3.2**

Sample size: 4096

| Parameter | Value | Probability | Parameter | Value | Probability | Parameter | Value | Probability |
|---|---|---|---|---|---|---|---|---|
| | | | $\bar{\kappa}$ = 0 | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.985 | $d$ | 0.25 | 0.861 | $d$ | 0.45 | 0.950 |
| | | | $\bar{\kappa}$ = 0 | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.870 | $d$ | 0.25 | 0.958 | $d$ | 0.45 | 0.944 |
| | | | $\bar{\kappa}$ = 0 | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.917 | $d$ | 0.25 | 0.943 | $d$ | 0.45 | 0.933 |
| | | | $\bar{\kappa}$ = 0 | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.984 | $d$ | 0.25 | 0.927 | $d$ | 0.45 | 0.921 |
| | | | $\bar{\kappa}$ = 1.5 | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.973 | $d$ | 0.25 | 0.980 | $d$ | 0.45 | 0.930 |
| | | | $\bar{\kappa}$ = 1.5 | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.893 | $d$ | 0.25 | 0.949 | $d$ | 0.45 | 0.941 |
| | | | $\bar{\kappa}$ = 1.5 | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.903 | $d$ | 0.25 | 0.921 | $d$ | 0.45 | 0.863 |
| | | | $\bar{\kappa}$ = 1.5 | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.974 | $d$ | 0.25 | 0.915 | $d$ | 0.45 | 0.862 |
| | | | $\bar{\kappa}$ = 3 | | $\phi_1$ = -0.8 | | | |
| $d$ | 0.15 | 0.974 | $d$ | 0.25 | 0.973 | $d$ | 0.45 | 0.873 |
| | | | $\bar{\kappa}$ = 3 | | $\phi_1$ = -0.4 | | | |
| $d$ | 0.15 | 0.909 | $d$ | 0.25 | 0.927 | $d$ | 0.45 | 0.872 |
| | | | $\bar{\kappa}$ = 1.5 | | $\phi_1$ = 0.4 | | | |
| $d$ | 0.15 | 0.971 | $d$ | 0.25 | 0.892 | $d$ | 0.45 | 0.840 |
| | | | $\bar{\kappa}$ = 3 | | $\phi_1$ = 0.8 | | | |
| $d$ | 0.15 | 0.972 | $d$ | 0.25 | 0.916 | $d$ | 0.45 | 0.891 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}(1-\phi_1 L)^{-1}$**. Design:** $\rho^2 = 1,\ \sigma^2 = 1,\ \mu = 0$**.**
**Theoretical probability: 0.95.**

**Table 4: Empirical size and power of LM test $\Phi$**

| | power | | | size | | |
|---|---|---|---|---|---|---|
| Values for $d$ | significance value | | | significance value | | |
| | 10 % | 5 % | 1 % | 10 % | 5 % | 1 % |
| | | | Sample size: 512 | | | |
| 0.1 | 0.143 | 0.093 | 0.046 | 0.079 | 0.047 | 0.016 |
| 0.2 | 0.370 | 0.298 | 0.184 | 0.096 | 0.054 | 0.012 |
| 0.375 | 0.602 | 0.541 | 0.455 | 0.073 | 0.036 | 0.018 |
| 0.45 | 0.613 | 0.556 | 0.471 | 0.080 | 0.044 | 0.018 |
| 0.49 | 0.461 | 0.402 | 0.313 | 0.078 | 0.037 | 0.015 |
| | | | Sample size: 1024 | | | |
| 0.1 | 0.184 | 0.124 | 0.066 | 0.082 | 0.052 | 0.014 |
| 0.2 | 0.424 | 0.343 | 0.214 | 0.084 | 0.041 | 0.009 |
| 0.375 | 0.692 | 0.647 | 0.541 | 0.094 | 0.045 | 0.010 |
| 0.45 | 0.673 | 0.625 | 0.534 | 0.092 | 0.048 | 0.016 |
| 0.49 | 0.529 | 0.475 | 0.380 | 0.095 | 0.056 | 0.017 |
| | | | Sample size: 2048 | | | |
| 0.1 | 0.184 | 0.134 | 0.055 | 0.093 | 0.050 | 0.010 |
| 0.2 | 0.457 | 0.371 | 0.249 | 0.082 | 0.045 | 0.006 |
| 0.375 | 0.744 | 0.708 | 0.621 | 0.089 | 0.049 | 0.010 |
| 0.45 | 0.750 | 0.703 | 0.631 | 0.088 | 0.052 | 0.008 |
| 0.49 | 0.589 | 0.515 | 0.422 | 0.112 | 0.052 | 0.020 |

**Parameterization:** $\alpha(L;\theta) = (1-L)^{-d}$. **Design:** $\rho^2 = 1$, $\sigma^2 = 1$, $\mu = 0$.

**Table 5.1: Empirical application of LM $\Phi$ test**

| | $S\&P\,500$ | $FTSE\,All$ | $FTSE\,100$ | $Yen/\$\,spot$ | $Yen/\$\,for$ | $\$/£\,spot$ | $\$/£\,for$ |
|---|---|---|---|---|---|---|---|
| $\Phi$ | 16460583 | 1440499 | 1229901 | 4.91 | 81.86 | 93.43 | 104.56 |
| $kurtosis$ | 100.26 | 27.78 | 26.15 | 6.69 | 7.52 | 7.17 | 7.18 |
| $skewness$ | -4.20 | -1.72 | -1.49 | -0.35 | -0.36 | 0.14 | 0.16 |

**LM test statistic $\Phi$, presented in first row, is asymptotically distributed like central $\chi^2_1$ under the null hypothesis of homoskedasticity.**

**Table 5.2: Semiparametric and Gaussian estimates of $d$**

| | $S\&P\,500$ | $FTSE\,All$ | $FTSE\,100$ | $Yen/\$\,spot$ | $Yen/\$\,1m\,for$ | $\$/£\,spot$ | $\$/£\,1m\,for$ |
|---|---|---|---|---|---|---|---|
| $semiparametric$ | 0.11 | 0.15 | 0.14 | 0.21 | 0.17 | 0.41 | 0.42 |
| $ARCH(\infty)$ | 0.33 | 0.35 | 0.37 | 0.32 | 0.24 | 0.54 | 0.56 |
| | (0.13) | (0.15) | (0.15) | (0.05) | (0.04) | (0.09) | (0.11) |
| $nonlinear\,MA$ | 0.35 | 0.36 | 0.43 | 0.30 | 0.26 | 0.41 | 0.42 |
| | (0.07) | (0.09) | (0.10) | (0.07) | (0.07) | (0.10) | (0.08) |
| | [3.06] | [3.13] | [4.07] | [3.58] | [2.25] | [1.74] | [1.81] |

**Row 1: values of local Gaussian estimator (Robinson 1995a), with bandwidth $m = 100$ and no trimming. The asymptotic standard error is equal to $1/(2\,m^{1/2}) = 0.050$.**
**Rows 2 and 3: values of Gaussian estimator pertinent to the values of $p, q$ $(0 \leq p, q \leq 4)$ yielding best goodness-of-fit of estimated residuals. Initial values given by grid search and semiparametric estimates of first row. Standard errors are in parentheses.**
**Estimates of $\bar{\kappa}$ are reported in square brackets.**

Table 5.3: Goodness-of-fit test of estimated residuals

| | ARCH($\infty$) | | nonlinear MA | | raw data | |
|---|---|---|---|---|---|---|
| | levels | squares | levels | squares | levels | squares |
| $S\&P500$ | 0.329 | 0.310 | 0.330 | 0.4047 | 0.319 | 0.196 |
| | [1.41] | [1.10] | [0.71] | [5.08] | [0.13] | [15.99] |
| $FTSE\,All$ | 0.342 | 0.239 | 0.346 | 0.350 | 0.333 | 0.366 |
| | [3.11] | [10.4] | [1.63] | [1.88] | [1.99] | [6.25] |
| $FTSE\,100$ | 0.307 | 0.168 | 0.336 | 0.363 | 0.321 | 0.370 |
| | [1.48] | [19.7] | [1.06] | [2.67] | [0.39] | [6.82] |
| $Yen/\$\,spot$ | 0.319 | 0.281 | 0.319 | 0.331 | 0.325 | 0.339 |
| | [0.15] | [4.83] | [0.05] | [0.76] | [0.80] | [2.72] |
| $Yen/\$\,for$ | 0.301 | 0.273 | 0.317 | 0.341 | 0.323 | 0.338 |
| | [2.27] | [5.95] | [-0.07] | [0.34] | [0.60] | [2.62] |
| $\$/£\,spot$ | 0.322 | 0.318 | 0.320 | 0.477 | 0.318 | 0.471 |
| | [0.46] | [0.07] | [0.11] | [9.32] | [0.01] | [20.1] |
| $\$/£\,for$ | 0.192 | 0.161 | 0.317 | 0.477 | 0.317 | 0.347 |
| | [16.5] | [20.6] | [-0.05] | [9.33] | [0.11] | [19.7] |

**The specification of Milhøj (1981) test for white-noise of sample of $z_t$ is**

$$W = \left(\frac{2\pi}{T}\sum_{s=1}^{T-1}(I_{zz}(\lambda_s))^2\right) / \left(\frac{2\pi}{T}\sum_{s=1}^{T-1} I_{zz}(\lambda_s)\right)^2.$$

**For each series, the $W$ test statistic, using the estimated residuals of best parameterization, in levels and squares, are reported. Under the null of perfect fit $T^{\frac{1}{2}}(W - 1/\pi)$ converges in distribution to $\mathcal{N}(0, 2/\pi^2)$. The studentized test in [ ]-brackets is asymptotically standard normal under the null of perfect fit.**

# References


Baillie, R. T., T. Bollerslev, and H. O. A. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74/1, 3–30.

Billingsley, P. (1986): *Probability and Measure*. New York: Wiley, second edn.

Black, F. (1976): "Studies of Stock Market Volatility Changes," in *Proceedings from the American Statistical Association, Business and Economic Statistics Section*.

Bollerlsev, T., R. Chou, and K. Kroner (1992): "ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

Bollerlsev, T., and H. Mikkelsen (1996): "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics*, 73/1, 151–184.

Bollerslev, T., R. Engle, and D. Nelson (1995): "ARCH Models," in *Handbook of Econometrics vol IV*, ed. by R. Engle,and D. McFadden. Amsterdam: North Holland.

Breidt, F., N. Crato, and P. deLima (1998): "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325–348.

Brillinger, D. (1975): *Time Series, Data Analysis and Theory*. New York: Holt Rinehart and Winston.

Brockwell, P., and R. Davis (1987): *Time Series: Theory and Methods*. New York: Springer Verlag.

Brown, B. (1971): "Martingale Central Limit Theorems," *The Annals of Mathematical Statistics*, 35, 1296–1303.

Campbell, J., and L. Hentschel (1990): "No News is Good News: an Asymmetric Model of Changing Volatility in Stock Returns," *Preprint*.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

Dahlhaus, R. (1989): "Efficient Parameter Estimation for Self-Similar Processes," *The Annals of Statistics*, 17, 1749–1782.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fama, E. (1965): "The Behaviour of Stock Market Prices," *Journal of Business*, 38, 34–105.

Fox, R., and M. Taqqu (1986): "Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *The Annals of Statistics*, 14/2, 517–532.

Fox, R., and M. Taqqu (1987): "Central Limit Theorems for Quadratic Forms in Random Variables having Long-Range Dependence," *Probability Theory and related Fields*, 74, 213–240.

Gallant, A., P. Rossi, and G. Tauchen (1993): "Nonlinear Dynamic Structure," *Econometrica*, 61/4, 871–907.

Giraitis, L., P. Kokoska, and R. Leipus (1998): "Stationary ARCH Models: Dependence Structure and Central Limit Theorem," *Econometric Theory*, 16, 3–22.

Giraitis, L., and P. M. Robinson (1998): "Least Squares (Whittle) Estimate Procedure for ARCH Models," *Preprint*.

Giraitis, L., and D. Surgailis (1990): "A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and its Application to Asymptotically Normality of Whittle's Estimate," *Probability Theory and Related Fields*, 86, 105–129.

Granger, C., and A. Andersen (1978): "On the Invertibility of Time Series Models," *Stochastic Processes and their Applications*, 8, 87–92.

Granger, C., and R. Joyeux (1980): "An Introduction to Long Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.

Grenander, U., and G. Szegö (1958): *Toeplitz Forms and their Applications*. Berkeley ; Los Angeles : University of California Press.

Hannan, E. (1970): *Multiple Time Series*. New York: Wiley.

Hannan, E. (1973): "The Asymptotic Theory of Linear Time Series Models," *Journal of Applied Probability*, 10, 130–145.

Hardy, G., J. Littlewood, and G. Polya (1964): *Inequalities.* Cambridge (UK): Cambridge University Press.

Harvey, A. (1998): "Long Memory in Stochastic Volatility," in *Forecasting volatility in financial markets*, ed. by J. Knight,and S. Satchell. London: Butterworth-Heineman, forthcoming.

Hentschel, L. (1995): "All in the Family: Nesting Symmetric and Asymmetric GARCH Models," *Journal of Financial Economics*, 39, 71–84.

Heyde, C., and R. Gay (1993): "Smoothed Periodogram Asymptotics and Estimation for Processes and Fields with Possible Long-Range Dependence," *Stochastic Processes and their Applications.*, 45, 169–182.

Hosoya, Y. (1993): "A Limit Theory on Stationary Processes with Long-Range Dependence and Related Statistical Models," Discussion paper 106, Faculty of Economics, Tohoku University.

Hosoya, Y. (1997): "A Limit Theory on Stationary Processes with Long-Range Dependence and Related Statistical Models," *The Annals of Statistics*, 25/1, 105–137.

Ibragimov, I., and J. Linnik (1971): *Independent and Stationary Sequences of Random Variables.* Gröningen: Walters-Noordhoff.

Lee, S., and B. Hansen (1994): "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Leonov, V., and A. Shiryaev (1959): "On a Method of Calculation of Semi-Invariants," *Theory of Probability Applications*, 4, 319–329.

Lintner, J. (1965): "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Stastistics*, 47, 13–37.

Lo, A. (1991): "Long Term Memory in Stock Market Prices," *Econometrica*, 59-5, 1279–1313.

Lumsdaine, R. (1996): "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," *Econometrica*, 64/3, 575–596.

Meddahi, N., and E. Renault (1997): "On Quadratic Estimating Functions for ARCH-Type Processes," *Preprint.*

Milhøj, A. (1981): "A Test of Fit in Time Series Models," *Biometrika*, 68-1, 177–187.

Monahan, J. (1984): "A Note on Enforcing Stationarity in Autoregressive-Moving Average Models," *Biometrika*, 71-2, 403–404.

Nelson, D. (1991): "Conditional Heteroscedasticity in Asset Pricing: a New Approach," *Econometrica*, 59, 347–370.

Robinson, P. M. (1977): "The Estimation of a Nonlinear Moving Average Model," *Stochastic Processes and their Applications*, 5, 81–90.

Robinson, P. M. (1978): "Alternative Models for Stationary Stochastic Processes," *Stochastic Processes and their Applications*, 8, 141–152.

Robinson, P. M. (1991): "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67–84.

Robinson, P. M. (1994a): "Rates of Convergence and Optimal Bandwidth for Long-Range Dependence," *Probability Theory and Related Fields*, 99, 443–473.

Robinson, P. M. (1994b): "Semiparametric Analysis of Long-Memory Time Series," *The Annals of Statistics*, 22-1, 515–539.

Robinson, P. M. (1995a): "Gaussian Semiparametric Estimation of Long Range Dependence," *Annals of Statistics*, 23, 1630–1661.

Robinson, P. M. (1995b): "Log-Periodogram Regression of Time Series with Long-Range Dependence," *Annals of Statistics*, 23, 1048–1072.

Robinson, P. M. (2001): "The Memory of Stochastic Volatility Models," *Journal of Econometrics*.

Robinson, P. M., and P. Zaffaroni (1997): "Modelling Nonlinearity and Long Memory in Time Series," in *Nonlinear Dynamics and Time Series, Fields Institute Communications, vol. 11*, ed. by C. Cutler,and D. Kaplan, pp. 161–170. Providence: American Mathematical Society.

Rozanov, Y. (1963): *Stationary Random Processes*. San Francisco: Holden Day.

Sharpe, W. (1964): "Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 1919, 425–442.

Stout, W. (1974): *Almost Sure Convergence*. New York; London: Academic Press Inc.

Taniguchi, M. (1982): "On Estimation of the Integrals of the Fourth Order Cumulant Spectral Density," *Biometrika*, 69, 117–122.

Taqqu, M. (1975): "Weak Convergence to Fractional Brownian Motion and the Rosenblatt Process," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31, 287–302.

Taylor, S. (1986): *Modelling Financial Time Series*. Chichester (UK): Wiley.

Weiss, A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 101–131.

Whittle, P. (1962): "Gaussian Estimation in Stationary Time Series," *Bulletin of the International Statistical Institute*, 39, 105–129.

Wright, J. (2000): "Log-Periodogram Estimation of Long-Memory Volatility Dependencies with Conditionally Heavy Tailed Returns," International Finance Discussion Papers 685, Board of Governors of the Federal Reserve System.

Yong, C. H. (1974): *Asymptotic Behaviour of Trigonometric Series*. Hong Kong: Chinese University of Hong Kong.

Zaffaroni, P. (1997): "Nonlinear long Memory Models with Applications in Finance," unpublished PhD thesis, LSE.

Zaffaroni, P. (2000): "Stationarity and Memory of ARCH($\infty$) Models," *Preprint*.

Zygmund, A. (1977): *Trigonometric Series*. Cambridge (UK): Cambridge University Press.

RECENTLY PUBLISHED "TEMI" (*)

No. 449 — *Contemporaneous aggregation of GARCH processes,* by P. ZAFFARONI (July 2002).

No. 450 — *Un'analisi critica delle definizioni di disoccupazione e partecipazione in Italia,* by E. VIVIANO (July 2002).

No. 451 — *Liquidity and announcement effects in the euro area,* by P. ANGELINI (October 2002).

No. 452 — *Misura e determinanti dell'agglomerazione spaziale nei comparti industriali in Italia,* by M. PAGNINI (October 2002).

No. 453 — *Labor market pooling: evidence from Italian industrial districts,* by G. DE BLASIO and S. DI ADDARIO (October 2002).

No. 454 — *Italian households' debt: determinants of demand and supply,* by S. MAGRI (October 2002).

No. 455 — *Heterogeneity in human capital and economic growth,* by S. ZOTTERI (October 2002).

No. 456 — *Real-time GDP forecasting in the euro area,* by A. BAFFIGI, R. GOLINELLI and G. PARIGI (December 2002).

No. 457 — *Monetary policy rules for the euro area: what role for national information?,* by P. ANGELINI, P. DEL GIOVANE, S. SIVIERO and D. TERLIZZESE (December 2002).

No. 458 — *The economic consequences of euro area modelling shortcuts,* by L. MONTEFORTE and S. SIVIERO (December 2002).

No. 459 — *Cross-country differences in self-employment rates: the role of institutions,* by R. TORRINI (December 2002).

No. 460 — *Dealing with forward-looking expectations and policy rules in quantifying the channels of transmission of monetary policy,* by F. ALTISSIMO, A. LOCARNO and S. SIVIERO (December 2002).

No. 461 — *Macroeconomics of international price discrimination,* by G. CORSETTI and L. DEDOLA (December 2002).

No. 462 — *Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth,* by G. D'ALESSIO and I. FAIELLA (December 2002).

No. 463 — *Metodologie di stima dell'economia sommersa: un'applicazione al caso italiano,* by R. ZIZZA (December 2002).

No. 464 — *Consolidation and efficiency in the financial sector: a review of the international evidence,* by D. AMEL, C. BARNES, F. PANETTA and C. SALLEO (December 2002).

No. 465 — *Human capital, technical change and the welfare state,* by R. BÉNABOU (December 2002).

No. 466 — *What do we learn from recall consumption data?,* by E. BATTISTIN, R. MINIACI and G. WEBER (February 2003).

No. 467 — *Evoluzione del sistema bancario e finanziamento dell'economia nel Mezzogiorno,* by F. PANETTA (March 2003).

No. 468 — *Transaction technology innovation and demand for overnight deposits in Italy,* by F. COLUMBA (March 2003).

No. 469 — *Sunk costs of exports,* by M. BUGAMELLI and L. INFANTE (March 2003).

No. 470 — *Testing against stochastic trend and seasonality in the presence of unattended breaks and unit roots,* by F. BUSETTI and A. M. R. TAYLOR (March 2003).

No. 471 — *Tax credit policy and firms' behaviour: the case of subsidies to open-end labour contracts in Italy, by* P. CIPOLLONE and A. GUELFI (April 2003).

---

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi – Divisione Biblioteca e pubblicazioni – Via Nazionale, 91 – 00184 Rome (fax 0039 06 47922059). They are available on the Internet at www.bancaditalia.it

"TEMI" LATER PUBLISHED ELSEWHERE

*1999*

L. GUISO and G. PARIGI, *Investment and demand uncertainty*, Quarterly Journal of Economics, Vol. 114 (1), pp. 185-228, **TD No. 289 (November 1996)**.

A. F. POZZOLO, *Gli effetti della liberalizzazione valutaria sulle transazioni finanziarie dell'Italia con l'estero*, Rivista di Politica Economica, Vol. 89 (3), pp. 45-76, **TD No. 296 (February 1997)**.

A. CUKIERMAN and F. LIPPI, *Central bank independence, centralization of wage bargaining, inflation and unemployment: theory and evidence*, European Economic Review, Vol. 43 (7), pp. 1395-1434, **TD No. 332 (April 1998)**.

P. CASELLI and R. RINALDI, *La politica fiscale nei paesi dell'Unione europea negli anni novanta,* Studi e note di economia, (1), pp. 71-109, **TD No. 334 (July 1998)**.

A. BRANDOLINI, *The distribution of personal income in post-war Italy: Source description, data quality, and the time pattern of income inequality*, Giornale degli economisti e Annali di economia, Vol. 58 (2), pp. 183-239, **TD No. 350 (April 1999)**.

L. GUISO, A. K. KASHYAP, F. PANETTA and D. TERLIZZESE, *Will a common European monetary policy have asymmetric effects?*, Economic Perspectives, Federal Reserve Bank of Chicago, Vol. 23 (4), pp. 56-75, **TD No. 384 (October 2000)**.

*2000*

P. ANGELINI, *Are Banks Risk-Averse? Timing of the Operations in the Interbank Market*, Journal of Money, Credit and Banking, Vol. 32 (1), pp. 54-73, **TD No. 266 (April 1996)**

F. DRUDI and R: GIORDANO, *Default Risk and optimal debt management,* Journal of Banking and Finance, Vol. 24 (6), pp. 861-892, **TD No. 278 (September 1996)**.

F. DRUDI and R. GIORDANO, *Wage indexation, employment and inflation,* Scandinavian Journal of Economics, Vol. 102 (4), pp. 645-668, **TD No. 292 (December 1996)**.

F. DRUDI and A. PRATI, *Signaling fiscal regime sustainability,* European Economic Review, Vol. 44 (10), pp. 1897-1930, **TD No. 335 (September 1998)**.

F. FORNARI and R. VIOLI, *The probability density function of interest rates implied in the price of options,* in: R. Violi, (ed.) , Mercati dei derivati, controllo monetario e stabilità finanziaria, Il Mulino, Bologna, **TD No. 339 (October 1998)**.

D. J. MARCHETTI and G. PARIGI, *Energy consumption, survey data and the prediction of industrial production in Italy,* Journal of Forecasting, Vol. 19 (5), pp. 419-440, **TD No. 342 (December 1998)**.

A. BAFFIGI, M. PAGNINI and F. QUINTILIANI, *Localismo bancario e distretti industriali: assetto dei mercati del credito e finanziamento degli investimenti,* in: L.F. Signorini (ed.), Lo sviluppo locale: un'indagine della Banca d'Italia sui distretti industriali, Donzelli, **TD No. 347** (**March 1999)**.

A. SCALIA and V. VACCA, *Does market transparency matter? A case study*, in: Market Liquidity: Research Findings and Selected Policy Implications, Basel, Bank for International Settlements, **TD No. 359 (October 1999)**.

F. SCHIVARDI, *Rigidità nel mercato del lavoro, disoccupazione e crescita*, Giornale degli economisti e Annali di economia, Vol. 59 (1), pp. 117-143, **TD No. 364 (December 1999)**.

G. BODO, R. GOLINELLI and G. PARIGI, *Forecasting industrial production in the euro area,* Empirical Economics, Vol. 25 (4), pp. 541-561, **TD No. 370 (March 2000)**.

F. ALTISSIMO, D. J. MARCHETTI and G. P. ONETO, *The Italian business cycle: Coincident and leading indicators and some stylized facts,* Giornale degli economisti e Annali di economia, Vol. 60 (2), pp. 147-220, **TD No. 377 (October 2000)**.

C. MICHELACCI and P. ZAFFARONI, *(Fractional) Beta convergence*, Journal of Monetary Economics, Vol. 45, pp. 129-153, **TD No. 383 (October 2000)**.

R. DE BONIS and A. FERRANDO, *The Italian banking structure in the nineties: testing the multimarket contact hypothesis*, Economic Notes, Vol. 29 (2), pp. 215-241, **TD No. 387 (October 2000)**.

*2001*

M. CARUSO, *Stock prices and money velocity: A multi-country analysis,* Empirical Economics, Vol. 26 (4), pp. 651-72, **TD No. 264 (February 1996)**.

P. CIPOLLONE and D. J. MARCHETTI, *Bottlenecks and limits to growth: A multisectoral analysis of Italian industry*, Journal of Policy Modeling, Vol. 23 (6), pp. 601-620, **TD No. 314 (August 1997)**.

P. CASELLI, *Fiscal consolidations under fixed exchange rates,* European Economic Review, Vol. 45 (3), pp. 425-450, **TD No. 336 (October 1998)**.

F. ALTISSIMO and G. L. VIOLANTE, *Nonlinear VAR: Some theory and an application to US GNP and unemployment,* Journal of Applied Econometrics, Vol. 16 (4), pp. 461-486, **TD No. 338 (October 1998)**.

F. NUCCI and A. F. POZZOLO, *Investment and the exchange rate*, European Economic Review, Vol. 45 (2), pp. 259-283, **TD No. 344 (December 1998)**.

L. GAMBACORTA, *On the institutional design of the European monetary union: Conservatism, stability pact and economic shocks,* Economic Notes, Vol. 30 (1), pp. 109-143, **TD No. 356 (June 1999)**.

P. FINALDI RUSSO and P. ROSSI, *Credit costraints in italian industrial districts,* Applied Economics, Vol. 33 (11), pp. 1469-1477, **TD No. 360 (December 1999)**.

A. CUKIERMAN and F. LIPPI, *Labor markets and monetary union: A strategic analysis,* Economic Journal, Vol. 111 (473), pp. 541-565, **TD No. 365 (February 2000)**.

G. PARIGI and S. SIVIERO, *An investment-function-based measure of capacity utilisation, potential output and utilised capacity in the Bank of Italy's quarterly model*, Economic Modelling, Vol. 18 (4), pp. 525-550, **TD No. 367 (February 2000)**.

F. BALASSONE and D. MONACELLI, *Emu fiscal rules: Is there a gap?,* in: M. Bordignon and D. Da Empoli (eds.), Politica fiscale, flessibilità dei mercati e crescita, Milano, Franco Angeli, **TD No. 375 (July 2000)**.

A. B. ATKINSON and A. BRANDOLINI, *Promise and pitfalls in the use of "secondary" data-sets: Income inequality in OECD countries*, Journal of Economic Literature, Vol. 39 (3), pp. 771-799, **TD No. 379 (October 2000)**.

D. FOCARELLI and A. F. POZZOLO, *The determinants of cross-border bank shareholdings: An analysis with bank-level data from OECD countries*, Journal of Banking and Finance, Vol. 25 (12), pp. 2305-2337, **TD No. 382 (October 2000)**.

M. SBRACIA and A. ZAGHINI, *Expectations and information in second generation currency crises models*, Economic Modelling, Vol. 18 (2), pp. 203-222, **TD No. 391 (December 2000)**.

F. FORNARI and A. MELE, *Recovering the probability density function of asset prices using GARCH as diffusion approximations,* Journal of Empirical Finance, Vol. 8 (1), pp. 83-110**, TD No. 396 (February 2001)**.

P. CIPOLLONE, *La convergenza dei salari manifatturieri in Europa,* Politica economica, Vol. 17 (1), pp. 97-125, **TD No. 398 (February 2001)**.

E. BONACCORSI DI PATTI and G. GOBBI, *The changing structure of local credit markets: Are small businesses special?*, Journal of Banking and Finance, Vol. 25 (12), pp. 2209-2237, **TD No. 404 (June 2001)**.

G. MESSINA, *Decentramento fiscale e perequazione regionale. Efficienza e redistribuzione nel nuovo sistema di finanziamento delle regioni a statuto ordinario*, Studi economici, Vol. 56 (73), pp. 131-148, **TD No. 416 (August 2001)**.

*2002*

R. CESARI and F. PANETTA, *Style, fees and performance of Italian equity funds*, Journal of Banking and Finance, Vol. 26 (1), **TD No. 325 (January 1998)**.

C. GIANNINI, *"Enemy of none but a common friend of all"? An international perspective on the lender-of-last-resort function*, Essay in International Finance, Vol. 214, Princeton, N. J., Princeton University Press, **TD No. 341 (December 1998)**.

A. ZAGHINI, *Fiscal adjustments and economic performing: A comparative study*, Applied Economics, Vol. 33 (5), pp. 613-624, **TD No. 355 (June 1999)**.

F. ALTISSIMO, S. SIVIERO and D. TERLIZZESE, *How deep are the deep parameters?,* Annales d'Economie et de Statistique,.(67/68), pp. 207-226, **TD No. 354 (June 1999)**.

F. FORNARI, C. MONTICELLI, M. PERICOLI and M. TIVEGNA, *The impact of news on the exchange rate of the lira and long-term interest rates,* Economic Modelling, Vol. 19 (4), pp. 611-639, **TD No. 358 (October 1999**).

D. FOCARELLI, F. PANETTA and C. SALLEO, *Why do banks merge?,* Journal of Money, Credit and Banking, Vol. 34 (4), pp. 1047-1066, **TD No. 361 (December 1999)**.

D. J. MARCHETTI, *Markup and the business cycle: Evidence from Italian manufacturing branches,* Open Economies Review, Vol. 13 (1), pp. 87-103, **TD No. 362 (December 1999**).

F. BUSETTI, *Testing for stochastic trends in series with structural breaks*, Journal of Forecasting, Vol. 21 (2), pp. 81-105, **TD No. 385 (October 2000)**.

F. LIPPI, *Revisiting the Case for a Populist Central Banker,* European Economic Review*,* Vol. 46 (3), pp. 601-612, **TD No. 386 (October 2000)**.

F. PANETTA, *The stability of the relation between the stock market and macroeconomic forces*, Economic Notes, Vol. 31 (3), **TD No. 393 (February 2001)**.

G. GRANDE and L. VENTURA, *Labor income and risky assets under market incompleteness: Evidence from Italian data,* Journal of Banking and Finance, Vol. 26 (2-3), pp. 597-620, **TD No. 399 (March 2001)**.

A. BRANDOLINI, P. CIPOLLONE and P. SESTITO, *Earnings dispersion, low pay and household poverty in Italy, 1977-1998,* in D. Cohen, T. Piketty and G. Saint-Paul (eds.), The Economics of Rising Inequalities, pp. 225-264, Oxford, Oxford University Press, **TD No. 427 (November 2001)**.

*2003*

F. SCHIVARDI, *Reallocation and learning over the business cycle,* European Economic Review, , Vol. 47 (1), pp. 95-111, **TD No. 345 (December 1998)**.

P. CASELLI, P. PAGANO and F. SCHIVARDI, *Uncertainty and slowdown of capital accumulation in Europe*, Applied Economics, Vol. 35 (1), pp. 79-89, **TD No. 372 (March 2000).**

E. GAIOTTI and A. GENERALE, *Does monetary policy have asymmetric effects? A look at the investment decisions of Italian firms,* Giornale degli Economisti e Annali di Economia, Vol. 61 (1), pp. 29-59, **TD No**. **429 (December 2001)**.

*FORTHCOMING*

A. F. POZZOLO, *Research and development regional spillovers, and the localisation of economic activities*, The Manchester School, **TD No. 331 (March 1998)**.

L. GAMBACORTA, *Asymmetric bank lending channels and ECB monetary policy*, Economic Modelling, **TD No. 340 (October 1998)**.

F. LIPPI, *Strategic monetary policy with non-atomistic wage-setters,* Review of Economic Studies, **TD No. 374 (June 2000)**.

P. ANGELINI and N. CETORELLI, *Bank competition and regulatory reform: The case of the Italian banking industry,* Journal of Money, Credit and Banking, **TD No. 380 (October 2000)**.

P. CHIADES and L. GAMBACORTA, *The Bernanke and Blinder model in an open economy: The Italian case,* German Economic Review, **TD No. 388 (December 2000)**.

P. PAGANO and F. SCHIVARDI, *Firm size distribution and growth,* Scandinavian Journal of Economics, **TD. No. 394 (February 2001)**.

M. PERICOLI and M. SBRACIA, *A Primer on Financial Contagion,* Journal of Economic Surveys, **TD No. 407 (June 2001)**.

M. SBRACIA and A. ZAGHINI, *The role of the banking system in the international transmission of shocks,* World Economy, **TD No. 409 (June 2001)**.

L. GAMBACORTA, *The Italian banking system and monetary policy transmission: Evidence from bank level data*, in: I. Angeloni, A. Kashyap and B. Mojon (eds.), Monetary Policy Transmission in the Euro Area, Cambridge, Cambridge University Press, **TD No. 430 (December 2001).**

M. EHRMANN, L. GAMBACORTA, J. MARTÍNEZ PAGÉS, P. SEVESTRE and A. WORMS, *Financial systems and the role of banks in monetary policy transmission in the euro area,* in: I. Angeloni, A. Kashyap and B. Mojon (eds.), Monetary Policy Transmission in the Euro Area, Cambridge, Cambridge University Press. **TD No. 432 (December 2001)**.

F. SPADAFORA, *Financial crises, moral hazard and the speciality of the international market: further evidence from the pricing of syndicated bank loans to emerging markets*, Emerging Markets Review, Vol. 4 ( 2), pp. 167-198, **TD No. 438 (March 2002)**.

D. FOCARELLI, *Bootstrap bias-correction procedure in estimating long-run relationships from dynamic panels, with an application to money demand in the euro area,* Economic Modelling, **TD No. 440 (March 2002)**.

D. FOCARELLI and F. PANETTA, *Are mergers beneficial to consumers? Evidence from the market for bank deposits*, American Economic Review, **TD No. 448 (July 2002)**.

A. BAFFIGI, R. GOLINELLI and G. PARIGI, *Bridge models to forecast the euro area GDP*, International Journal of Forecasting, **TD No. 456 (December 2002)**.

F. BUSETTI and A. M. ROBERT TAYLOR, *Testing against stochastic trend and seasonality in the presence of unattended breaks and unit roots*, Journal of Econometrics, **TD No. 470 (February 2003)**.